

Beiträge zur Numerik der Parameterschätzung
bei räumlichen Interaktionsmodellen

von
Ludwig Götz
aus
Marktrechwitz

1980 № 289



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41546219_0077

Beiträge zur Numerik der Parameterschätzung bei räumlichen Interaktionsmodellen

Den Naturwissenschaftlichen Fakultäten
der Friedrich-Alexander-Universität Erlangen-Nürnberg

zur

Erlangung des Doktorgrades

vorgelegt

von

Ludwig Götz

aus

Marktredwitz



Als Dissertation genehmigt von den Naturwissenschaft-
lichen Fakultäten der Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 8.2.1980

Vorsitzender der Promotionskommission:

Professor Dr. H. Wegener

Erstberichterstatter: Professor Dr. B. Dejon

Zweitberichterstatter: Professor Dr. D. Kölzow

INHALTSVERZEICHNIS

Einleitung	i
§1. Allgemeine Vorbemerkungen	i
§2. Überblick über die Arbeit	x
 Kapitel I: Gravitationsmodelle - Darstellung, theoretische Begründung und Kalibrierungsverfahren	 1
§1. Mathematische Verkehrstheorie	1
§2. Gravitationsmodelle	2
2.1. Verkehrsverteilung als Gleichgewichtsmodell in einem Netzwerk	3
2.2. Die Entropie-Maximierungsmethode	8
2.3. Verallgemeinerte Kosten	13
§3. Kalibrierung von Verkehrsverteilungsmodellen	15
3.1. Kalibrierung als "inverses Problem"	15
3.2. Kalibrierungsverfahren für Gravitationsmodelle	16
§4. Kalibrierung von Cobb-Douglas-Funktionen	23
 Kapitel II: Approximation in der l_1 - Norm	 29
§1. Der einparametrische Fall: $K=1$	31
1.1. Untersuchung des Spezialfalles $M=3$ mit äquidistanten x -Werten	34
1.2. Unimodalitätskriterien	40
1.2.1. Sturmsche Ketten	40
1.2.2. Residuensatz	43
1.2.3. Abschätzungen	44
§2. Der mehrparametrische Fall	49
 Kapitel III: Approximation in der l_2 - Norm	 54
§1. Der einparametrische Fall: $K=1$	57
§2. Der mehrparametrische Fall	62
§3. Fixpunktiteration und verallgemeinerte nicht-lineare Eigenwertmethoden	67
§4. Lineare l_2 -Approximation logarithmierter Werte	77

Kapitel IV: Approximation in der l_∞ -Norm	83
§1. Konvexität der Minimalmenge	85
§2. Numerische Näherungsverfahren zur Lösung	90
Kapitel V: Anwendung der mathematischen Methoden auf das Verkehrsverteilungsmodell	96
§1. Vorbemerkungen	96
§2. Unimodalitätsuntersuchungen	97
§3. Optimierungsverfahren für das l_2 -Kriterium	99
§4. Optimierungsverfahren für das l_∞ -Kriterium	107
Bezeichnungen	110
Literaturverzeichnis	112

EINLEITUNG:

§1. Allgemeine Vorbemerkungen

Gegenstand der vorliegenden Arbeit sind Beiträge zur Numerik der Parameterschätzung bei räumlichen Interaktionsmodellen. Wir beschäftigen uns dabei insbesondere mit der Kalibrierung von Gravitationsmodellen, die in der mathematischen Verkehrstheorie zur Darstellung der Verkehrsverteilung dienen.

Gravitationsmodelle beschreiben die Zahl der Fahrten T_{ij} von einer Quellzone i in eine Zielzone j als proportional zu einem Produkt von Exponentialfunktionen der sogenannten erklärenden Variablen $x_{ij}^{(k)}$ (z.B. Reisezeit, Entfernung, Fahrtkosten):

$$T_{ij} = l_i k_j \prod_{k=1}^K e^{\beta_k x_{ij}^{(k)}}, \quad \begin{matrix} i=1, \dots, I, \\ j=1, \dots, J. \end{matrix}$$

Dabei sind l_i , k_j und β_k zu bestimmende Parameter. Zu gegebenen β_k erhält man die l_i und k_j aus Konsistenzbedingungen der Art nämlich

$$\sum_{j=1}^J T_{ij} = \hat{O}_i, \quad i=1, \dots, I,$$

$$\sum_{i=1}^I T_{ij} = \hat{D}_j, \quad j=1, \dots, J,$$

wobei $\hat{O}_i$ bzw. $\hat{D}_j$ die empirischen Werte aller von einer Quellzone i ausgehenden bzw. in eine Zielzone j einmündenden Fahrten sind.

Unter der Kalibrierung eines Modells versteht man die Bestimmung der Parameter derart, daß die mit ihnen berechneten Modellergebnisse die empirischen Daten bezüglich eines Gütekriteriums "möglichst gut" approximieren.

Deutet man die Quell- und Zielzonen als Knoten eines mathematischen Netzwerks, dann läßt sich das Verkehrsverteilungsproblem als Gleichgewichtsproblem in diesem Netzwerk formulieren. Gibt man bestimmte Kennlinien für die Bögen dieses Netzes vor (vgl. Kap. I), so erhält man als Gleichung für die Komponenten des Gleichgewichtsstromes

$$T_{ij} = e^{\lambda_i - x_j} \prod_{k=1}^K e^{\beta_k} x_{ij}^{(k)}$$

mit den Knotenpotentialen λ_i und x_j , die die natürlichen Logarithmen der Parameter l_i und $1/k_j$ sind. Die obigen Konsistenzbedingungen stellen dann die Kirchhoffschen Knotenbedingungen für den Gleichgewichtsstrom dar.

Um das in der Verkehrsliteratur häufig verwendete einfachere Modell

$$T_{ij} = l_i k_j e^{-\beta c_{ij}}$$

mit nur einer einzigen erklärenden Variablen c_{ij} , den sogenannten verallgemeinerten Reisekosten, zu kalibrieren, setzt man meist alle $\hat{O}_i$, $\hat{D}_j$ und c_{ij} als bekannt voraus und bestimmt die Parameter l_i und k_j so, daß die Konsistenzbedingungen gelten. Der Parameter β wird so festgelegt, daß die mittleren verallgemeinerten Reisekosten des Modells mit dem empirischen

Wert übereinstimmen, daß also die Bedingung

$$\frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij}}{\sum_{i=1}^I \sum_{j=1}^J p_{ij}} = \frac{\sum_{i=1}^I \sum_{j=1}^J c_{ij} \hat{T}_{ij}}{\sum_{i=1}^I \sum_{j=1}^J \hat{T}_{ij}}$$

erfüllt wird ([14] [17] [18] [44] [45]).

Dieses Kriterium zur Kalibrierung von β ergibt sich aus der Begründung der Gravitationsmodelle mit Hilfe der Entropie-Maximierungsmethode (siehe Kap.I). Wenn zur Kalibrierung des Modells statt der $\hat{O}_i$ und $\hat{D}_j$ die empirischen Daten $\hat{T}_{ij}$ für die Zahl der Fahrten verwendet werden, kann man auch andere Kriterien zur Bestimmung des Parameters β heranziehen, etwa l_p -Normen für die Differenz der aus den T_{ij} und $\hat{T}_{ij}$ gebildeten Vektoren $\mathbf{t}$ und $\hat{\mathbf{t}}$.

Wir untersuchen in dieser Arbeit als Gütekriterien für die Kalibrierung der Parameter β_k die l_1 -, die l_2 - und die l_∞ -Norm, da sich die l_1 -Norm als Summe der insgesamt fehlverteilten Fahrten interpretieren läßt, die l_2 -Norm (stellvertretend für andere l_p -Normen) wegen ihrer Differenzierbarkeit und die l_∞ -Norm, um den maximalen (oder bei Rechnung auf Prozentbasis den maximalen prozentualen) Fehler zu minimieren.

Da die empirischen Daten im allgemeinen aus statistischen Erhebungen und Verkehrszählungen stammen, sind sie mit Meßfehlern behaftet. Ebenso werden die empirischen Daten $\hat{T}_{ij}$ auch

normalerweise nicht genau einem Gleichgewichtszustand des Modellnetzes entsprechen, sei es, weil bei der Wahl der Modellstruktur zu starke Vereinfachungen in Hinblick auf die Zahl der eingehenden erklärenden Variablen $x_{ij}^{(k)}$ und damit der zugehörigen Parameter β_k getroffen wurden, sei es, daß man statt des vorliegenden statischen Modells ein dynamisches benötigte. Wir setzen zur Kalibrierung des Modells voraus, daß die Matrix $\hat{T} := (\hat{T}_{ij})$ der empirischen Fahrten zur Verfügung steht.

Allerdings gewinnen wir aus diesen Daten dann nicht die $\hat{O}_i$ und $\hat{D}_j$, um mit Hilfe der Parameter l_i und k_j mittels eines Furness- oder Knotenanalysealgorithmus (vgl. Kap. I) die exakte Erfüllung der Konsistenzbedingungen zu erzwingen und dabei die $\hat{T}_{ij}$, die detailliertere Information enthalten, mit Hilfe der β_k nur näherungsweise zu approximieren. Stattdessen verwenden wir sämtliche in die Modellgleichung eingehenden Parameter dazu, eines der Gütekriterien

$$\|t - \hat{t}\|_p, \quad p = 1, 2, \infty,$$

zu minimieren (vgl. [31]).

Je besser die theoretischen Werte T_{ij} die empirischen $\hat{T}_{ij}$ approximieren, desto besser approximieren auch die über die Konsistenzbedingungen abgeleiteten O_i und D_j die entsprechenden empirischen Werte $\hat{O}_i$ und $\hat{D}_j$. Durch die angedeutete Vorgehensweise erhält man ein unrestringiertes Optimierungsproblem, das jedoch höherdimensional wird und statt K nun $K+I+J$ Variable enthält.

Da sich in der Ökonometrie bei der Kalibrierung von Cobb-Douglas-Funktionen ein ähnliches unrestringiertes nichtlinea-

res Optimierungsproblem stellt, gehen wir kurz auf die dort aufgestellten Hypothesen über die Verteilung der Meßfehler und die Optimierungsverfahren ein, die sich bei Verwendung von Maximum-Likelihood-Methoden ergeben.

Der von uns unter anderen untersuchte Fall eines von der l_2 - Norm abgeleiteten Gütekriteriums läßt sich dabei auf die Parameterschätzung bei Cobb-Douglas-Funktionen anwenden, wenn die Meßfehler additiv und normalverteilt sind. Schätzt man die Parameter der Cobb-Douglas-Funktionen mit Hilfe der exponentiellen Regression, dann erhält man auf Grund der linearen Struktur der logarithmierten Modellgleichung mit der Methode der kleinsten Fehlerquadrate rasch Lösungen. Sind die Meßfehler dabei multiplikativ und log-normalverteilt, dann sind die so gewonnenen Lösungen Maximum-Likelihood-Lösungen.

Wir geben ein Verfahren an, das durch Gewichtung der Fehlergleichungen in erster Näherung den Fehler ausgleicht, den man macht, wenn man bei Vorliegen additiver normalverteilter Meßfehler exponentielle Regression zur Parameterschätzung anwendet. Auf diese Weise erhalten wir schnell relativ gute Ausgangslösungen für die l_2 - Approximation (Kap.III, §4).

In der Approximationstheorie wird unter anderem die Frage nach der Existenz und der Eindeutigkeit einer besten Approximierenden bei der diskreten l_p - und der kontinuierlichen L_p - Approximation durch Exponentialsummen

$$L(A, x) := \sum_{n=1}^N a_n e^{\beta_n x},$$

$$A := (a_1, \dots, a_N, \beta_1, \dots, \beta_N),$$

und bei der kontinuierlichen L_D -Approximation mit Summen von Produkten von Exponentialfunktionen ([28][30])

$$g(A, x) := \sum_{n=1}^N a_n \prod_{k=1}^K e^{\beta_n^k x^{(k)}},$$

$$A := (a_1, \dots, a_N, \beta_1^1, \dots, \beta_1^K, \dots, \beta_N^1, \dots, \beta_N^K),$$

$$x := (x^{(1)}, \dots, x^{(K)}),$$

behandelt. Wir geben in den entsprechenden Kapiteln Beispiele dafür an, daß sich die Ergebnisse nicht auf unser Problem der diskreten Approximation in den $I \cdot J$ Punkten x_{ij} mit

$$f(\pi, x_{ij}) := l_i k_j \prod_{k=1}^K e^{\beta_k x_{i,j}^{(k)}},$$

$$\pi := (l_1, \dots, l_I, k_1, \dots, k_J, \beta_1, \dots, \beta_K),$$

übertragen lassen.

Was den Spezialfall der Bestimmung des Parameters β im Ansatz

$$T_{ij} = l_i k_j e^{-\beta c_{ij}}$$

bei bekannten Parametern l_i und k_j angeht, so liegt es nahe zu versuchen, ihn der Approximation mit Exponentialsummen mit $N=1$ unterzuordnen. Dies ist nicht möglich, da in unserem Problem der Parameter a fehlt. Wir werden für den Fall der l_1 -

und der L_2 -Approximation Beispiele angeben (siehe Kap. II und Kap. III), bei denen entweder die beste Approximierende der Form

$$L(A, x) = a e^{\beta x} \quad \text{mit } A = (a, \beta)$$

eindeutig ist, während es zwei beste Approximierende der Form

$$f(\beta, x) = e^{\beta x}$$

gibt, oder bei denen zwei beste Approximierende der Form $L(A, x)$ auftreten und die Approximierende $f(\beta, x)$ eindeutig ist.

Anhand von Beispielen zeigen wir, daß es nicht möglich ist, globale Eindeutigkeitsaussagen für die L_n -Approximation ($n=1,2$) zu treffen. (Zu diesem Fragenkreis siehe auch [6][7].) Auch kann man daraus, daß zu einer Funktion auf einem Intervall eine eindeutige beste L_n -Approximierende existiert, nicht darauf schließen, daß dann auch stets die diskrete L_n -Approximierende zu einer Diskretisierung dieser Funktion eindeutig wäre. Wir werden dies in Kap. III an einem Beispiel veranschaulichen.

Insgesamt lassen die von uns behandelten Beispiele vermuten, daß dann Eindeutigkeit der besten Approximierenden vorliegt, wenn die zu approximierenden Punkte bereits auf einer Exponentialfunktion liegen oder die Abweichungen davon gering sind, d.h. die Meßfehler nicht zu groß werden.

Im Hinblick auf die Verwendung von Abstiegsverfahren zur Be-

rechnung der besten Approximierenden ist die Unimodalität des Approximationsfehlers in Abhängigkeit von den zu bestimmenden Parametern nützlich. Die Eindeutigkeit des Minimums des Approximationsfehlers genügt nicht, da sie noch lokale Nebenminima zuläßt und Abstiegsverfahren gegen ein solches konvergieren könnten.

Deshalb geben wir hinreichende Kriterien an, mit deren Hilfe man für den einparametrischen Fall der Bestimmung des Parameters β feststellen kann, falls Unimodalität vorliegt ($p=1, 2$). Für die diskrete Tschebyscheff-Approximation mit Produkten von Exponentialfunktionen beweisen wir für den mehrparametrischen Fall unter sehr geringen Einschränkungen, die in der Praxis im allgemeinen erfüllt sein dürften, die Unimodalität und damit auch die Eindeutigkeit der l_∞ -besten Approximierenden.

Außer mit Unimodalitätsuntersuchungen beschäftigt sich diese Arbeit auch mit Berechnungsverfahren für die optimalen Parameter.

So geben wir für den Fall der l_2 -Approximation ein dem Problem angepaßtes Verfahren an, das durch Blockiteration abwechselnd mit Hilfe von Fixpunktiteration die Parameter l_i und k_j und mit Hilfe eines modifizierten Newton-Verfahrens die Parameter β_k bestimmt, wobei jeweils die anderen Parameter festgehalten werden.

Um die l_∞ -optimalen Parameter zu berechnen, formulieren wir das nichtlineare Problem in eine äquivalente Folge von linearen Problemen um, bei denen jeweils mittels eines

Simplex-Programms über die Verträglichkeit eines Systems von linearen Ungleichungen entschieden wird.

Ein Vergleich der von uns verwendeten l_n -Kriterien und Methoden zur Kalibrierung mit den sonst üblichen zeigt folgendes: Ein Nachteil des l_1 - und des l_2 -Kriteriums ist die Tatsache, daß man Beispiele konstruieren kann, für die das Optimum nicht eindeutig ist, was zur Divergenz oder zum Oszillieren der Algorithmen führen kann oder dazu, daß das Verfahren nicht das globale, sondern nur ein lokales Optimum liefert. Beim l_∞ -Kriterium treten diese Probleme nicht auf, dafür können hier bekanntlich einzelne große Meßfehler die Bestimmung der optimalen Parameter stärker beeinflussen als beim l_1 - oder l_2 -Kriterium.

Verwendet man die Parameter l_i und k_j zusammen mit den β_k zur Optimierung des Gütekriteriums, so werden im allgemeinen die Konsistenzbedingungen verletzt. Allerdings scheint diese Verletzung gerade in gewissen "kritischen" Knoten besonders groß zu sein, da ein Vergleich mit einem anderen Verfahren, bei dem die l_i und k_j zur exakten Erfüllung der Konsistenzbedingungen dienten und nur die β_k zur Optimierung des Gütekriteriums verwendet wurden, zeigte, daß dort die theoretischen Werte für die Zahl der Fahrten auf den eben von diesen Knoten ausgehenden bzw. in sie einmündenden Bögen die empirischen Werte besonders schlecht approximierten. Hier scheint also das Modell die Wirklichkeit nicht gut zu beschreiben.

Im Fall der l_2 -Approximation liefert das von uns verwendete

Verfahren zur Kalibrierung doppelt so schnell eine Lösung, als wenn man die Parameter l_i und k_j zur Erfüllung der Konsistenzbedingungen und nur die β_k zur Optimierung des Gütekriteriums verwendet.

Zusammenfassend läßt sich sagen, daß keine der beiden Vorgehensweisen deutliche Vor- oder Nachteile der andern gegenüber hat. Wenn es darum geht, möglichst schnell relativ gute Ausgangswerte für die Parameter zu bekommen, dann sollte man alle in die Modellgleichung eingehenden Parameter dazu verwenden, das Gütekriterium zu optimieren.

§2. Überblick über die Arbeit

Die vorliegende Arbeit gliedert sich in fünf Kapitel. Im ersten geben wir die allgemeinen Voraussetzungen und Definitionen für die nachfolgenden an. Dabei stellen wir eine theoretische Begründung des Gravitationsmodells dar und machen einige Anmerkungen zur sogenannten Entropie-Maximierungsmethode. Ferner beschreiben wir kurz die Cobb-Douglas-Funktionen, ohne sie allerdings theoretisch zu begründen, sowie die Probleme, die bei bestimmten Hypothesen über die Verteilung der Meßfehler bei Verwendung von Maximum-Likelihood-Schätzfunktionen auftreten.

Kapitel II beschäftigt sich mit der l_1 -Approximation des Vektors $\hat{\mathbf{t}}$ der empirischen Fahrten $\hat{t}_{ij}$ durch den Vektor $\mathbf{t}$ der

theoretischen Fahrten T_{ij} :

$$\|t(\pi, x) - \hat{t}\|_1 = \sum_{i=1}^I \sum_{j=1}^J |T_{ij}(\pi, x) - \hat{T}_{ij}|.$$

Dabei ist x der Vektor der erklärenden Variablen und π der Parametervektor, der außer den K Parametern β_k auch noch die $I+J$ Potentiale $\lambda_i = \ln l_i$ und $\chi_j = -\ln k_j$ enthält. Das obige Gütekriterium ist keine konvexe Funktion von π . Ebenso braucht nicht für beliebige zu approximierende Punkte $(x_{ij}, \hat{T}_{ij})$ Unimodalität des Gütekriteriums bezüglich π vorzuliegen. Wir geben aber wenigstens für den Fall nur eines zu kalibrierenden Parameters β mit Hilfe der Taylor-Entwicklung des Gütekriteriums hinreichende Abschätzungen 0., 1. und 2. Ordnung an, mit denen man feststellen kann, ob die l_1 -Fehlerfunktion unimodal ist. Weitere hinreichende Unimodalitätskriterien entwickeln wir aus dem Residuensatz und den Sturmschen Ketten. Dieser eindimensionale Spezialfall tritt beispielsweise dann auf, wenn entweder in einem Blockiterationsverfahren abwechselnd die Potentiale und der Parameter β optimiert werden, oder wenn die Potentiale mit Hilfe eines Knotenanalysealgorithmus zur Erfüllung der Konsistenzbedingungen verwendet werden und nur ein Parameter β zur Optimierung des Gütekriteriums dient.

Im Kapitel III betrachten wir die Approximation in der l_2 -Norm, also

$$\|t(\pi, x) - \hat{t}\|_2.$$

Dazu untersuchen wir das Gütekriterium

$$\phi(\pi) := \sum_{i=1}^I \sum_{j=1}^J (T_{ij}(\pi, x_{ij}) - \hat{T}_{ij})^2,$$

das wie bei der l_1 -Norm weder eine konvexe noch eine für beliebige Paare $(x_{ij}, \hat{T}_{ij})$ unimodale Funktion von π darstellt. Auch hier geben wir für den eindimensionalen Fall hinreichende Unimodalitätskriterien an, die auf Residuensatz und Sturmschen Ketten beruhen.

Für den in der Verkehrs- und Wirtschaftstheorie eigentlich interessierenden Fall mehrerer zu kalibrierender Parameter untersuchen wir ein Verfahren, das unter der Voraussetzung der Unimodalität gegen das Minimum des Gütekriteriums konvergiert. Der besonderen Struktur des Gravitationsmodells, nämlich, daß in dem Ansatz

$$T_{ij} = e^{\lambda_i - x_j} \prod_{k=1}^K e^{\beta_k x_{ij}^{(k)}}$$

vor den Potentialen λ_i und x_j nur jeweils +1 oder -1 als Faktor steht, während bei den Parametern β_k im allgemeinen von ± 1 verschiedene erklärende Variable stehen können, tragen wir dadurch Rechnung, daß wir die optimalen Potentiale und Parameter durch ein Blockiterationsverfahren bestimmen, bei dem die Potentiale durch Fixpunktiteration, die Parameter durch ein modifiziertes Newtonverfahren abwechselnd optimiert werden. Auf diese Weise umgehen wir bei der Bestimmung der Potentiale die Berechnung der inversen Hessematrix, die nötig wäre, wollte man das Gesamtproblem mittels des Newtonverfahrens lösen.

Die nähere Analyse des Fixpunktverfahrens führt auf ein verallgemeinertes nichtlineares Eigenwertproblem. Für den Fall, daß die Potentiale bezüglich des folgenden, von einer gewichteten l_2 -Norm abgeleiteten Gütekriteriums

$$\sum_{i=1}^I \sum_{j=1}^J \frac{1}{\sigma_{ij}^2} (\hat{T}_{ij} - e^{\lambda_i - x_j} \sigma_{ij})^2$$

mit den Gewichten $\frac{1}{\sigma_{ij}^2}$,

$$\sigma_{ij} : = \prod_{k=1}^K e^{\beta_k x_{ij}^{(k)}},$$

berechnet werden, wird das Eigenwertproblem linear.

Vir zeigen, daß dann e^{λ_i} und e^{-x_j} die Komponenten des Eigenvektors zum größten Eigenwert sind, der bei einer positiven Matrix nach dem Satz von Perron einfach ist, woraus die Eindeutigkeit des Minimums des Gütekriteriums folgt.

Schließlich enthält das dritte Kapitel noch eine Untersuchung des Fehlers, den man macht, wenn man als Näherungslösung für das Optimierungsproblem

$$\min \phi(\pi)$$

das Minimum des Problems

$$\min \psi(\pi)$$

mit

$$\psi(\pi) := \sum_{i=1}^I \sum_{j=1}^J (\ln T_{ij}(x_{ij}, \pi) - \ln \hat{T}_{ij})^2$$

benutzt. Da $\ln T_{ij}(x_{ij}, \pi)$ linear von den Potentialen und Parametern abhängt, kann man rasch das Minimum von ψ angeben, indem man das überbestimmte lineare Gleichungssystem

$$\lambda_i - x_j + \sum_{k=1}^K \beta_k x_{ij}^{(k)} = \ln \hat{T}_{ij}, \quad \begin{matrix} i=1, \dots, I, \\ j=1, \dots, J, \end{matrix}$$

nach der Methode der kleinsten Fehlerquadrate löst. Die Verschiebung der Minima von ϕ und ψ gegeneinander läßt sich in erster Näherung rückgängig machen, indem man die Gleichungen mit T_{ij} wichtet. Man nützt so die Schnelligkeit des Fehler-

quadratverfahrens aus und erhält trotzdem bessere Näherungswerte für das Minimum von ϕ als mit dem ungewichteten Verfahren.

Kapitel IV behandelt die Approximation in der Tschebyscheff-Norm, also das Optimierungsproblem

$$\min_{\pi} \left(\max_{i,j} \left\{ \left| T_{ij}(\pi, x_{ij}) - \hat{T}_{ij} \right| \right\} \right).$$

Wir beweisen zunächst die unimodale Abhängigkeit der Fehlerfunktion von den Parametern und lösen dann dieses nichtlineare Optimierungsproblem durch eine Folge von linearen Programmen.

In Kapitel V folgt die Anwendung der theoretischen Überlegungen auf die Kalibrierung des Verkehrsverteilungsmodells anhand von Daten aus dem Gesamtverkehrsplan der Stadt Erlangen. Wir geben die Ergebnisse der Kalibrierungsverfahren an und machen einige technische Bemerkungen zu den Algorithmen.

Um triviale Fälle auszuschließen, setzen wir in der Arbeit grundsätzlich voraus, daß die Vektoren der erklärenden Variablen stets den gesamten $\mathbb{R}^K$ aufspannen und mindestens so viele zu approximierende Punkte $(x_{ij}, \hat{T}_{ij})$ wie Parameter existieren.

Ich danke Herrn Prof. Dr. B. Dejon für die Anregungen und Ratschläge, mit denen er diese Arbeit unterstützt hat.

KAPITEL I:

GRAVITATIONSMODELLE - DARSTELLUNG, THEORETISCHE BEGRÜNDUNG UND KALIBRIERUNGSVERFAHREN

Dieses Kapitel stellt zunächst die Gravitationsmodelle der Verkehrstheorie als Gleichgewichtsproblem in einem mathematischen Netzwerk dar. Es folgen Anmerkungen zur theoretischen Begründung durch die sogenannte Entropie-Maximierungsmethode. Schließlich stellen wir noch die Cobb-Douglas-Funktionen und die statistischen Methoden zu ihrer Parameterschätzung, die in der Ökonometrie unter bestimmten Voraussetzungen über die Meßfehlerverteilung üblich sind, vor.

§1. Mathematische Verkehrstheorie

In der mathematischen Verkehrstheorie werden Modelle für das Verkehrsgeschehen erstellt, um mit deren Hilfe den gegenwärtigen Verkehr zu analysieren und Prognosen über seine zukünftige Entwicklung abgeben zu können. Zur Beschreibung des Stadtverkehrs verwendet man ein internettes Gleichgewichtsmodell, dessen Lösung man in vier Phasen berechnen kann, nämlich

- Verkehrserzeugung
- Verkehrsverteilung
- Modalsplit
- Verkehrswegwahl.

Allerdings ist die Trennung zwischen diesen Teilmodellen nicht scharf, teilweise werden Verkehrsverteilung und Modalsplit oder Verkehrsverteilung und -wegwahl auch zusammengefaßt.

In der Verkehrserzeugung werden aus Strukturdaten und Planungsdaten die Zahlen O_i und D_j der Fahrten bestimmt, die von einer Zone i ausgehen oder in eine Zone j einmünden. Damit wird in der Verkehrsverteilung die Pendler- oder Fahrtenmatrix $T = (T_{ij})$ errechnet, die angibt, wieviele Pendler T_{ij} von einer bestimmten Zone i einer Stadt in eine andere Zone j fahren.

Das Modalsplitmodell untersucht, wie sich die Pendler auf die zwischen zwei Zonen zur Verfügung stehenden Verkehrsmittel verteilen, und bei der Verkehrswegewahl wird der Verkehr der verschiedenen Verkehrsmittelnetze auf die möglichen Wege von i nach j umgelegt.

Verkehrsverteilungsmodelle beruhen auf drei Grundannahmen:

Die Zahl der Fahrten T_{ij} zwischen zwei Zonen i und j ist

- direkt proportional zur Zahl O_i der insgesamt von der Quellzone i ausgehenden Fahrten,
- direkt proportional zur Zahl D_j der insgesamt in die Zielzone j einmündenden Fahrten,
- verringert sich mit zunehmendem Aufwand, der erbracht werden muß, um von i nach j zu kommen.

§2. Das Gravitationsmodell

Ein häufig verwendetes Modell zur Verkehrsverteilung, das diesen drei Forderungen genügt, ist das Gravitationsmodell. Sein Name rührt daher, daß eine der frühesten Formulierungen [37] analog zum Newtonschen Gravitationsgesetz aufgebaut war:

$$T_{ij} = \gamma \frac{O_i D_j}{c_{ij}^2} . \quad (I.2.1)$$

γ ist dabei ein Proportionalitätsfaktor.

Allgemeiner versteht man heute unter Gravitationsmodell einen Ansatz der Form

$$T_{ij} = l_i k_j f(x_{ij}, \beta) \quad (I.2.2)$$

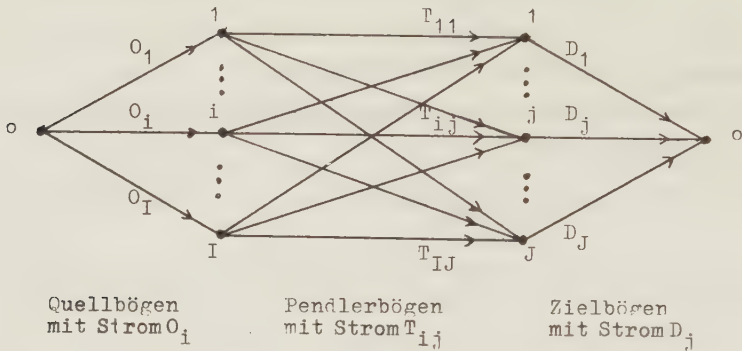
mit quell- und zielzonenabhängigen Parametern l_i und k_j und einer Funktion f , die abnimmt, wenn der Aufwand, der nötig ist, um von i nach j zu kommen, wächst. x_{ij} bezeichnet dabei, wie schon weiter oben angedeutet, den Vektor der erklärenden Variablen.

2.1. Verkehrsverteilung als Gleichgewichtsmodell in einem Netzwerk

Die mathematische Netzwerktheorie bietet eine Möglichkeit zur theoretischen Begründung des Gravitationsmodells, die im Vergleich zu dem Ansatz von Murchland [34] und der Entropie-Maximierungsmethode von Wilson [44] den Vorteil hat, auch auf andere, kompliziertere Netztopologien anwendbar zu sein. Im Verkehrsverteilungsmodell teilt man die Stadt in I Quell- und J Zielzonen ein, die durch die Knoten des Netzes repräsentiert werden. Die zwischen diesen Zonen herrschenden Pendlerbeziehungen stellt man durch den Fluß auf den Bögen dar, die die Knoten verbinden.

Dann bildet ein Netzwerk folgender Gestalt ein Modell für die Verkehrsverteilung:

Abbildung 1: Quellzonen mit Knotenpotential λ_i Zielzonen mit Knotenpotential π_j



Wir stellen zunächst einige Definitionen aus der Netzwerktheorie zusammen, die wir im folgenden benötigen.

I.1. Definition:

Die Struktur des Netzes wird durch die sogenannte Knoteninzidenzmatrix D zum Ausdruck gebracht, deren Zeilenanzahl gleich der Zahl der Knoten und deren Spaltenanzahl gleich der Zahl der Bögen des Netzes ist. Für das Element d_{kl} von D gilt:

$$d_{kl} := \begin{cases} +1, & \text{falls Bogen } l \text{ von Knoten } k \text{ abgeht} \\ -1, & \text{falls Bogen } l \text{ in Knoten } k \text{ einmündet} \\ 0 & \text{sonst} \end{cases}$$

I.2. Definition:

Sei K die Menge der Knoten und B die Menge der Bögen.

Eine Knotenbelegung des Netzes (K, B) ist eine Abbildung

$$p : K \longrightarrow \mathbb{R}.$$

Eine Bogenbelegung des Netzes (K, B) ist eine Abbildung

$$q : B \longrightarrow \mathbb{R}.$$

Wir schreiben diese Abbildungen als Vektoren $p \in \mathbb{R}^{|K|}$ und $g \in \mathbb{R}^{|B|}$.

Eine Bogenbelegung $f \in \mathbb{R}^{|B|}$ heißt Strom (Zirkulation), wenn sie im Nullraum $\mathcal{N}(\mathbb{D})$ der Knoteninzidenzmatrix $\mathbb{D}$ liegt, wenn also

$$\mathbb{D} f = 0. \quad (\text{I.2.3})$$

Die Gleichung (I.2.3) heißt auch Kirchhoffsche Knotenbedingung für Ströme.

Eine Bogenbelegung $t \in \mathbb{R}^B$ heißt Spannung (Tension), wenn sie im Bildraum $\mathcal{B}(\mathbb{D}^t)$ der zu $\mathbb{D}$ transponierten Matrix $\mathbb{D}^t$ liegt. Das bedeutet, daß dann eine Knotenbelegung $p \in \mathbb{R}^{|K|}$ existiert mit

$$\mathbb{D}^t p = t.$$

p heißt dann Vektor der Knotenpotentiale, und die Komponenten der Spannung ergeben sich für jeden Bogen als Differenzen der Potentiale in den Endpunkten des Bogens.

I.3. Bemerkung:

In unserem Fall gilt $|B| = I + IJ + J$. Fordert man bei der Verkehrsverteilung die Erfüllung der Konsistenzbedingungen (siehe Einleitung Seite -i-), also

$$\sum_{j=1}^J T_{ij} = 0_i, \quad i = 1, \dots, I, \quad (\text{I.2.4})$$

$$\sum_{i=1}^I T_{ij} = D_j, \quad j = 1, \dots, J \quad (\text{I.2.5})$$

und damit auch

$$\sum_{i=1}^I 0_i = \sum_{j=1}^J D_j, \quad (\text{I.2.6})$$

so genügt der Vektor

$$f := (0_1, 0_2, \dots, 0_I, T_{11}, T_{12}, \dots, T_{IJ}, D_1, D_2, \dots, D_J) \in \mathbb{R}^{|B|}$$

den Kirchhoffschen Knotenbedingungen für Ströme. f ist daher ein Strom.

Der Zustand eines Netzwerks wird durch zwei Zustandsvariable, nämlich einen Strom $f \in \mathbb{R}^{|B|}$ und eine Spannung $t \in \mathbb{R}^{|B|}$ beschrieben. Außerdem liegt je Bogen b ein Konstitutivgesetz vor, das die Komponenten f_b und t_b miteinander verknüpft. Dieses Gesetz wird graphisch durch die sogenannte Kennlinie $\mathcal{K}_b$ dargestellt.

I.4. Definition:

Eine isotone Kennlinie $\mathcal{K}_b$ eines Bogens $b \in B$ in der (f_b, t_b) -Ebene ist das Bild eines offenen Intervalls $(\underline{x}, \bar{x})$ mit $-\infty \leq \underline{x} \leq \bar{x} \leq +\infty$ unter einer stetigen Abbildung $(f_b(x), t_b(x))$ mit $x \in (\underline{x}, \bar{x})$ und den Eigenschaften:

$$i) \quad (f_b(x_1) - f_b(x_2)) (t_b(x_1) - t_b(x_2)) \geq 0 \quad (I.2.7)$$

$$ii) \quad \lim_{x \rightarrow \underline{x}} (|f_b(x)| + |t_b(x)|) = \lim_{x \rightarrow \bar{x}} (|f_b(x)| + |t_b(x)|) = +\infty \quad (I.2.8)$$

Liegt $(\hat{f}_b, \hat{t}_b)$ auf $\mathcal{K}_b$, so heißt $(\hat{f}_b, \hat{t}_b)$ konstitutiv zulässig.

Schreibweise: $f_b = k^{-1}(t_b)$, $t_b = k(f_b)$ (k und k^{-1} können mengenwertig sein).

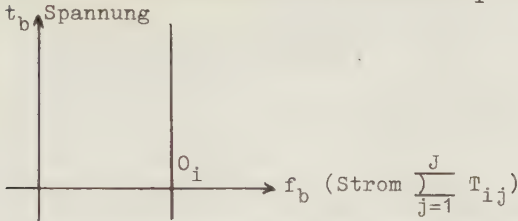
I.5. Definition:

Das geordnete Paar (f, t) von Bogenbelegungen heißt Gleichgewichtszustand für das Netzwerk $(K, B, \mathcal{K})$ mit einer Menge $\mathcal{K} := \{\mathcal{K}_b / b \in B\}$ isotoner Kennlinien, wenn gilt:

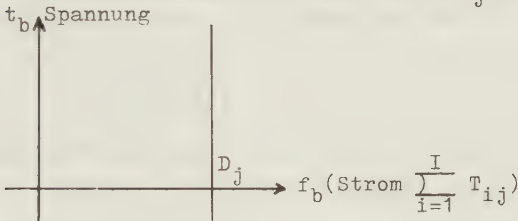
- i) f ist ein Strom.
- ii) t ist eine Spannung.
- iii) f und t sind konstitutiv zulässig: $(f_b, t_b) \in \mathcal{K}_b \quad \forall b \in B$.

Man erhält das Gravitationsmodell zur Verkehrsverteilung, indem man den Vektor f als Gleichgewichtsstrom in dem obigen Netzwerk mit nachfolgenden Kennlinien postuliert:

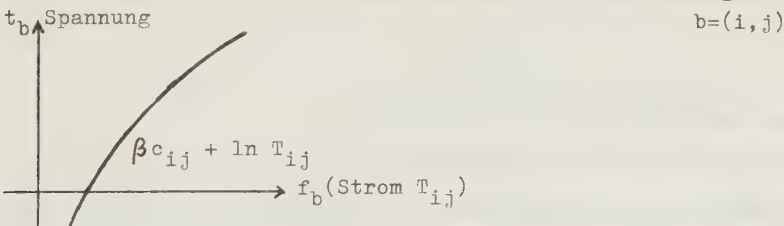
a) Ideale Stromquellen der Stärke O_i auf den Quellbögen $b=(o,i)$



b) Ideale Stromquellen der Stärke D_j auf den Zielbögen $b=(j,o)$



c) Sogenannte Gravitationskennlinien auf den Pendlerbögen



Die sich auf den Pendlerbögen einstellende Spannung besteht aus einem den Kosten c_{ij} proportionalen Term und einem, der logarithmisch mit dem Strom wächst.

Die Formeln (ii) und (iii) besagen, daß es für jeden Quellknoten i und für jeden Zielknoten j Potentialwerte λ_i und χ_j gibt, so daß auf jedem Pendlerbogen (ij) gilt:

$$\lambda_i - \chi_j = \beta c_{ij} + \ln T_{ij}. \quad (I.2.9)$$

Lösen nach T_{ij} liefert dann die Formeln des Gravitationsmodells

$$T_{ij} = e^{\lambda_i - \chi_j} e^{-\beta c_{ij}} \quad (\text{I.2.10})$$

bzw.

$$T_{ij} = l_j k_j e^{-\beta c_{ij}} \quad (\text{I.2.11})$$

zu denen - wegen (i) - noch die Konsistenzbedingungen (I.2.4-6) hinzukommen. Da in die Gleichung (I.2.9) nur die Potentialdifferenzen eingehen, sind die Potentiale bis auf eine additive Konstante bestimmt. Setzt man jedoch das Potential im "Referenzknoten" fest (i.a. $\chi_j = 0$), dann sind die Potentiale eindeutig bestimmt.

2.2. Die Entropie-Maximierungsmethode

Eine andere Möglichkeit zur Begründung des Gravitationsmodells zur Verkehrsverteilung bildet die Entropie-Maximierungsmethode von Wilson [44]. Da sich ein Großteil der Verkehrsliteratur auf sie beruft, wollen wir diese Methode aus Gründen der Abgeschlossenheit dieser Arbeit kurz darstellen.

Wir gehen wieder von dem in Abbildung 1 dargestellten Verkehrsverteilungsproblem aus. Für jede Quellzone i sei die Zahl O_i der insgesamt von ihr ausgehenden Fahrten und für jede Zielzone j die Zahl D_j der insgesamt in sie einmündenden Fahrten bekannt, außerdem die sogenannten mittleren Reisekosten c:

$$c = \frac{1}{T} \frac{\sum_{i=1}^I}{\sum_{j=1}^J} c_{ij} T_{ij} \quad (\text{I.2.12})$$

$$\text{mit} \quad T = \sum_{i=1}^I \sum_{j=1}^J T_{ij} \quad (\text{I.2.13})$$

oder die Gesamtreisekosten

$$C = \sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij} . \quad (\text{I.2.14})$$

Wilson betrachtet drei Beschreibungsniveaus für das vorliegende System, die sich durch die in ihnen enthaltene Information unterscheiden:

- (O_i, D_j, C) - Niveau,
- (T) - Niveau,
- das sogenannte Niveau der Mikrozustände.

Am detailliertesten ist die Beschreibung des Systems auf dem Niveau der Mikrozustände, bei dem für jede einzelne Fahrt Quell- und Zielzone bekannt sind. Von diesem Niveau aus lassen sich die anderen beiden eindeutig bestimmen.

Das mittlere Niveau ist durch die Fahrtenmatrix gekennzeichnet, die die Gesamtzahlen T_{ij} der Fahrten von jeder Quell- in jede Zielzone enthält. Zwar kann man daraus noch eindeutig das (O_i, D_j, C) - Niveau bestimmen, aber es gibt viele Möglichkeiten, durch Mikrozustände ein festes (T) - Niveau zu realisieren.

Das (O_i, D_j, C) - Niveau wird schließlich nur noch durch die Größen O_i , D_j und C charakterisiert, und es gibt viele Möglichkeiten, es durch (T) - Niveaus zu realisieren.

Wilson setzt nun voraus, daß alle Mikrozustände, die eine bestimmte Konfiguration T ergeben, gleichwahrscheinlich sind.

Die Wahrscheinlichkeit, daß ein bestimmtes T dann wirklich auftritt, ist proportional zu der Zahl $W(T)$ der Möglichkeiten, T durch Mikrozustände zu realisieren.

Es gibt
$$W(\overline{T}) := \frac{T!}{\prod_{i,j} T_{ij}!} \quad (\text{I.2.15})$$

Möglichkeiten, T Fahrten so in Gruppen zu T_{ij} zusammenzufassen, daß die Konfiguration $\overline{T}$ entsteht. Den wahrscheinlichsten mit dem (O_i, D_j, C) -Niveau verträglichen Zustand $\overline{T}$ erhält man durch Maximierung von $W(\overline{T})$ bzw. von

$$S := \ln W(\overline{T}) \quad (\text{I.2.16})$$

unter den Restriktionen

$$O_i = \sum_{j=1}^J T_{ij}, \quad i = 1, \dots, I, \quad (\text{I.2.4})$$

$$D_j = \sum_{i=1}^I T_{ij}, \quad j = 1, \dots, J, \quad (\text{I.2.5})$$

$$C = \sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij}. \quad (\text{I.2.14})$$

Hierzu ersetzt man mit Hilfe der Stirlingschen Formel

$$\ln(N!) \approx N \ln N - N + \frac{1}{2} \ln N + \ln 2 \quad (\text{I.2.17})$$

und ihrer Ableitung

$$\frac{d \ln(N!)}{dN} \approx \ln N + \frac{1}{2N} \approx \ln N \text{ für große } N \quad (\text{I.2.18})$$

(I.2.16) durch den einfacheren Näherungsausdruck

$$\begin{aligned} S &= T \ln T - \sum_{i=1}^I \sum_{j=1}^J T_{ij} \ln T_{ij} = \\ &= \sum_{i=1}^I \sum_{j=1}^J T_{ij} (\ln T - \ln T_{ij}). \end{aligned} \quad (\text{I.2.19})$$

Die Lage des Maximums bestimmt man mit Hilfe der Lagrange-

Technik, indem man den Gradienten der Lagrange-Funktion

$$\begin{aligned}
 L_S(T, \lambda, \beta) : = & S + \beta \left(C - \sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij} \right) \\
 & + \sum_{i=1}^I \lambda_i \left(\sum_{j=1}^J T_{ij} - O_i \right) \\
 & + \sum_{j=1}^J \alpha_j \left(D_j - \sum_{i=1}^I T_{ij} \right) \quad (I.2.20)
 \end{aligned}$$

gleich Null setzt. Als Bedingungen für ein Maximum erhält man

$$T_{ij} = l_i k_j e^{-\beta c_{ij}} \quad (I.2.11)$$

mit

$$l_i := e^{\lambda_i} = \frac{1}{\sum_{j=1}^J k_j e^{-\beta c_{ij}}} \quad (I.2.21)$$

und

$$k_j := e^{-\alpha_j} = \frac{1}{\sum_{i=1}^I l_i e^{-\beta c_{ij}}} \quad (I.2.22)$$

I.6. Bemerkung:

Das hier beschriebene Verfahren verdankt den Namen Entropie-Maximierungsmethode dem Term S in (I.2.10), der analog zur Entropie in der statistischen Mechanik oder in der Informationstheorie definiert ist.

Ersetzt man T_{ij} durch

$$p_{ij} := \frac{T_{ij}}{T}, \quad (I.2.23)$$

dann folgt $\sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1$ und

$$S := - \sum_{i=1}^I \sum_{j=1}^J p_{ij} \ln p_{ij}, \quad (I.2.24)$$

also die in der statistischen Mechanik (vgl.[20]) und in der Informationstheorie (vgl.[26]) übliche Definition der Entropie.

I.7. Bemerkung:

Auf ähnliche Weise wie Wilson gewinnt Murchland [34] das Gravitationsmodell. Während Wilson die Entropie unter den Nebenbedingungen (I.2.4), (I.2.5) und (I.2.14) maximiert, minimiert Murchland ein gewichtetes Mittel aus negativer Entropie und Gesamtreisekosten (I.2.14) unter den Nebenbedingungen (I.2.10) und (I.2.11):

$$\min_{\mathbf{T}} M(\mathbf{T})$$

$$M(\mathbf{T}) := \sum_{i=1}^I \sum_{j=1}^J T_{ij} \ln T_{ij} + \beta \sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij} \quad (\text{I.2.25})$$

Erländer und Stewart [15] schließlich betrachten das Verkehrsverteilungsproblem als Minimalkostenproblem und verlangen die Minimierung der Gesamtreisekosten (I.2.14) unter einer Nebenbedingung an die Entropie und unter (I.2.4) und (I.2.5).

$$\min C(\mathbf{T})$$

$$C(\mathbf{T}) := \sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij} \quad (\text{I.2.14})$$

unter

$$S(\mathbf{T}) := - \sum_{i=1}^I \sum_{j=1}^J T_{ij} \ln T_{ij} \geq S_0 \quad (\text{I.2.26})$$

wobei S_0 als empirische Größe als bekannt vorausgesetzt wird.

In beiden Fällen führt die Bestimmung eines stationären Punktes der zugehörigen Lagrange-Funktion wieder auf die Gleichung (I.2.11).

2.3 Verallgemeinerte Kosten

Bei den bisherigen Betrachtungen hatte die Funktion der erklärenden Variablen stets die Form

$$f(c_{ij}) = e^{-\beta c_{ij}}. \quad (I.2.27)$$

Man kann diesen Ansatz jedoch leicht verallgemeinern, indem man bei der netzwerktheoretischen Darstellungsweise eine Kennlinie benützt, die zusätzliche erklärende Variablen berücksichtigt, oder beim Entropie-Ansatz zusätzliche Restriktionen vorschreibt.

Allgemein kann die Funktion f das folgende Aussehen haben:

$$f_g(x_{ij}) := \prod_{m=1}^M (x_{ij}^{om})^{\alpha_g^m} e^{\beta_g^m \cdot x_{ij}^{om}} \prod_{l=1}^L \prod_{m=1}^M e^{\beta_g^{lm} \cdot x_{ij}^{lm}} \prod_{r=1}^R e^{\gamma_g^r \cdot x_i^r} \prod_{s=1}^S e^{\delta_g^s \cdot x_j^s}. \quad (I.2.28)$$

In diesem Fall tritt an die Stelle des Parameters β aus (I.2.27) der Parametervektor β

$$\beta_g := (\alpha_g^1, \dots, \alpha_g^M, \beta_g^1, \dots, \beta_g^M, \beta_g^{11}, \dots, \beta_g^{LM}, \gamma_g^1, \dots, \gamma_g^R, \delta_g^1, \dots, \delta_g^S)^t$$

und an die Stelle der erklärenden Variablen c_{ij} der Vektor von erklärenden Variablen

$$x_{ij} := (\ln x_{ij}^{o1}, \dots, \ln x_{ij}^{oM}, x_{ij}^{o1}, \dots, x_{ij}^{oM}, x_{ij}^{11}, \dots, x_{ij}^{LM}, x_i^1, \dots, x_i^R, x_j^1, \dots, x_j^S)^t$$

Es werden also drei Typen von Variablen unterschieden: bogenbezogene mit ij indizierte (z.b. x_{ij}^{om} Reisezeit von i nach j)

im Netz m , x_{ij}^{1m} Verfügbarkeit im Netz m , x_{ij}^{2m} Barauslasten etc.), quellzonenbezogene mit i indizierte (z.B. x_i^1 mittleres Einkommen in Zone i , x_i^2 Fahrzeuge pro Familie, etc.) und zielzonenbezogene mit j indizierte erklärende Variable (z.B. Zahl der Arbeitsplätze, Zahl der Wohnungen etc.).

Daß die Reisezeit x_{ij}^{om} in der Form

$$(x_{ij}^{om})^{\alpha_g^m} e^{\beta_g^m x_{ij}^{om}}$$

in die Funktion f eingeht, trägt dem Rechnung, daß im allgemeinen für kleine Entfernungen keine Verkehrsmittel verwendet werden und man folgende Abhängigkeit von der Reisezeit erwartet:

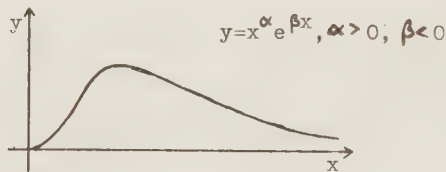


Abb.2:

Die Indizierung bedeutet:

- $i = 1, \dots, I$: Quellzone,
- $j = 1, \dots, J$: Zielzone,
- $m = 1, \dots, M$: Verkehrsmittelnetz,
- $l = 1, \dots, L$: bogenbezogene Variable,
- $r = 1, \dots, R$: quellzonenbezogene Variable,
- $s = 1, \dots, S$: zielzonenbezogene Variable.

Der Index g steht für die sogenannte Gruppeneinteilung, d.h. dafür, daß jeweils für bestimmte Gruppen von Stadtzonen (z.B. solche mit gleicher Kfz-Dichte oder gleichem Haushaltseinkommen) bestimmte Parameter gelten. Läßt man g weg, so bedeutet das, daß keine Gruppeneinteilung getroffen wird und die Parameter für alle Zonen gelten.

Weitere Ansätze des Typs (I.2.28) findet man in der Literatur [23] .

§3 Kalibrierung von Verkehrsverteilungsmodellen

3.1. Kalibrierung als "inverses Problem"

Während man beim Gleichgewichtsproblem in Netzwerken Kennlinien und Flüsse auf den Quell- und Zielbögen als bekannt voraussetzt und danach fragt, welche konstitutiv zulässigen Ströme und Spannungen sich einstellen, hat man es bei der Kalibrierung mit einem inversen Problem zu tun:

Unter der Voraussetzung, daß das Verkehrsverteilungsproblem durch ein Gleichgewicht beschrieben wird, gibt man die Netzstruktur und eine ein- oder mehrparametrische Familie von Kennlinien vor und bestimmt die unbekannten Parameter so, daß die sich damit ergebenden Ströme und Potentiale konstitutiv zulässig sind und die Modellergebnisse die empirischen Daten bezüglich eines Gütekriteriums möglichst gut approximieren.

Dabei tritt das Problem auf, daß man bei Verwendung empirischer Daten, die etwa durch Haushaltsbefragungen gewonnen werden oder sich aus Planungsdaten ergeben, mit Meßfehlern rechnen muß, so daß diese Daten nur annähernd einem Gleichgewicht entsprechen. Außerdem können Abweichungen vom Gleichgewicht auch darauf beruhen, daß man es im allgemeinen wohl mit einem nicht-stationären System zu tun hat und das statische Modell die Wirklichkeit nur unvollkommen wiedergibt.

3.2. Kalibrierungsverfahren für Gravitationsmodelle

Je nachdem, welche Gütekriterien Verwendung finden, welche Restriktionen zu erfüllen sind und welche Parameter kalibriert werden sollen, unterscheiden wir drei Fälle:

a) Bestimme $\beta = \beta^*$ in dem Ansatz

$$T_{ij}(\lambda(\beta), \beta, x_{ij}) = e^{\lambda_i - x_j} f(\beta, x_{ij}) \quad (I.3.1)$$

$$\text{mit } \lambda := (\lambda_1, \dots, \lambda_I, x_1, \dots, x_J)^t$$

derart, daß gilt:

$$\hat{O}_i = \sum_{j=1}^J T_{ij}(\lambda(\beta^*), \beta^*, x_{ij}), \quad i=1, \dots, I, \quad (I.3.2)$$

$$\hat{D}_j = \sum_{i=1}^I T_{ij}(\lambda(\beta^*), \beta^*, x_{ij}), \quad j=1, \dots, J, \quad (I.3.3)$$

$$\hat{C} = \sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij}(\lambda(\beta^*), \beta^*, x_{ij}). \quad (I.3.4)$$

Die Schreibweise $\lambda = \lambda(\beta)$ erklärt sich daraus, daß auf Grund der Restriktionen λ eindeutig von β abhängt (vgl. 3.2.1).

Bemerkung: Es sollen alle Restriktionen (I.3.2-4) erfüllt werden, wobei es genausoviele Restriktionen gibt wie Parameter und Potentiale zusammen. Dabei wird nicht berücksichtigt, wie gut die empirischen Fahrtenzahlen $\hat{T}_{ij}$ approximiert werden. Ihre Kenntnis ist zur Kalibrierung des Modells nur indirekt erforderlich, nämlich zur Berechnung von $\hat{O}_i$, $\hat{D}_j$ und $\hat{C}$. Fall (a) tritt vor allem in der Literatur auf, die sich auf die Entropie-Maximierungsmethode beruft ([16][7][8][44][45]).

Das Modell enthält nur den Parameter β , jedoch lassen sich durch Hinzufügen weiterer Restriktionen vom Typ (I.3.4) noch zusätzliche erklärende Variable und damit Parameter einbauen. Von den Restriktionen (I.3.2) und (I.3.3) sind je $I+J-1$ linear

unabhängig, so daß man mit ihrer Hilfe auch nur $I+J-1$ Potentiale bestimmen kann. Dem entspricht die Tatsache, daß man das Potential in einem Referenzknoten beliebig festlegen kann, weil nur die Potentialdifferenzen in die Modellgleichung eingehen.

b) Bestimme β^* in dem Ansatz

$$T_{ij}(\lambda(\beta), \beta, x_{ij}) = e^{\lambda_i - x_j} f(\beta, x_{ij}) \quad (I.3.5)$$

derart, daß gilt:

$$\hat{O}_i = \sum_{j=1}^J T_{ij}(\lambda(\beta^*), \beta^*, x_{ij}), \quad i=1, \dots, I, \quad (I.3.6)$$

$$\hat{D}_j = \sum_{i=1}^I T_{ij}(\lambda(\beta^*), \beta^*, x_{ij}), \quad j=1, \dots, J, \quad (I.3.7)$$

$$\|t(\lambda(\beta^*), \beta^*, x) - \hat{t}\| \leq \|t(\lambda(\beta), \beta, x) - \hat{t}\|, \quad \forall \beta \in R^M, \quad (I.3.8)$$

wobei $\|\cdot\|$ eine der beliebigen Normen im R^{IJ} und $t(\lambda(\beta), \beta, x)$ den Vektor der $T_{ij}(\lambda(\beta), \beta, x_{ij})$ (in lexikographischer Reihenfolge) bezeichnet.

Bemerkung: Zwar wird immer noch verlangt, daß die Zeilen- und Spaltensummen aus empirischer und theoretischer Fahrtenmatrix exakt übereinstimmen, zusätzlich sollen jetzt aber auch die einzelnen empirischen Fahrtenanzahlen durch das Modell möglichst genau approximiert werden.

c) Bestimme λ^* und β^* in dem Ansatz

$$T_{ij}(\lambda, \beta, x_{ij}) = e^{\lambda_i - x_j} f(\beta, x_{ij}) \quad (I.3.9)$$

derart, daß gilt:

$$\|t(\lambda^*, \beta^*, x) - \hat{t}\| \leq \|t(\lambda, \beta, x) - \hat{t}\|, \beta \in \mathbb{R}^M, \lambda \in \mathbb{R}^{I+J}. \quad (I.3.10)$$

Bemerkung: Hier sind keine Restriktionen mehr zu erfüllen.

Das Gütekriterium wird in Abhängigkeit von sämtlichen Potentialen und Parametern optimiert. Während in den Fällen (a) und (b) die Erfüllung der Restriktionen (I.3.2) und (I.3.3) bzw. (I.3.6) und (I.3.7) äquivalent dazu ist, daß die Kirchhoffschen Knotenbedingungen (I.2.3) gelten und damit

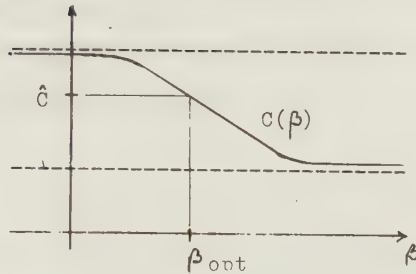
$$(\hat{O}_1, \dots, \hat{O}_I, \hat{T}_{11}, \dots, \hat{T}_{IJ}, \hat{D}_1, \dots, \hat{D}_J)$$

ein Strom in dem von uns betrachteten Verkehrsverteilungsnetzwerk ist, gilt dies im Fall (c) im allgemeinen nicht mehr. Ebenso sind jetzt die Potentiale λ und die Parameter β "entkoppelt", d.h. die Potentiale sind jetzt keine Funktion der Parameter mehr, wie es vorher die zu den Strömen T_{ij} gehörigen konstitutiv zulässigen Gleichgewichtspotentiale waren. Wenn jedoch die theoretischen T_{ij} die empirischen $\hat{T}_{ij}$ gut approximieren, dann werden auch die Restriktionen (I.3.6) und (I.3.7) gut erfüllt (vgl. [10] [31]).

3.2.1 Kalibrierung im Fall (a)

Die Kalibrierung beruht hier auf zwei Grundüberlegungen. Einerseits liegen auf Grund der Knotenbedingungen die Potentiale bei bekanntem Parameter β und vorgegebenen $\hat{O}_i$ und $\hat{D}_j$ fest (wir berücksichtigen dies durch die Schreibweise $\lambda(\beta)$), andererseits bilden die mit den Gleichgewichtsströmen ebenfalls gegebenen Gesamtreisekosten eine in β antitone Funktion. Diese Funktion hat den in Abbildung 3 dargestellten Verlauf.

Abb. 3:
(vgl. [18])



Bestimmung des Parameters β

Für den Fall (a) sind uns zwei Verfahren bekannt, die aus der Funktion $C(\beta)$ der Gesamtreisekosten im Gleichgewicht den Parameter β so bestimmen, daß theoretische und empirische Gesamtreisekosten übereinstimmen.

Das eine Verfahren [18] geht von einem beliebigen Anfangswert für β aus, berechnet dazu die Potentiale und Ströme und gewinnt daraus die zugehörigen Gesamtreisekosten. Diese werden mit dem empirischen Wert verglichen und darauf der Parameter β vergrößert bzw. verkleinert, je nachdem ob $C(\beta) > \hat{C}$ oder $C(\beta) < \hat{C}$ ist. In jedem Schritt muß dabei ein Knotenanalyseverfahren durchlaufen werden, um Gleichgewichtspotentiale und -ströme zu berechnen.

Dies ist bei dem zweiten Verfahren [17] nicht nötig, das auf der Taylor-Entwicklung der Funktion $C(\beta)$ beruht. Allerdings führt diese Methode rasch zu komplizierten Ausdrücken, wenn man Taylor-Polynome höherer Ordnung (z.B. $N=10$) zur Approximation von $C(\beta)$ verwendet, da jeweils die $(n+1)$ -te Ableitung von C auf die n -te ($1 \leq n \leq N-1$) zurückgeführt werden muß, was für jede Stufe die Lösung eines Gleichungssystems $(I+J)$ -ter Ordnung erfordert.

Bestimmung der Potentiale λ_i und x_j (bei gewähltem β)

Sei β fest. Dann ist die Funktion $f(\beta, c_{ij})$ bekannt und aus den Restriktionen (I.3.2) und (I.3.3) erhält man die Potentiale mit Hilfe des im folgenden beschriebenen Knotenanalyseverfahrens.

I.8. Algorithmus: (Knotenanalyse allgemein)

1. Durchlaufe zyklisch alle Knoten und berechne die Ergiebigkeit E , die folgendermaßen definiert ist:

$$\text{im Knoten } i: \quad E(i) := \sum_{j=1}^J k_{ij}^{-1} (\lambda_i - x_j) - \hat{O}_i \quad (\text{I.3.11})$$

$$\text{im Knoten } j: \quad E(j) := \hat{D}_j - \sum_{i=1}^I k_{ij}^{-1} (\lambda_i - x_j) \quad (\text{I.3.12})$$

(k_{ij}^{-1} ist dabei die Inverse zu der die Kennlinie $\mathcal{K}_{ij}$ definierenden Funktion k_{ij})

2. Falls die Ergiebigkeit in einem Knoten kleiner Null ist, vergrößere das entsprechende Potential (λ_i oder x_j), bis sie Null wird.

Falls sie größer Null ist, verkleinere das entsprechende Knotenpotential, bis sie Null wird.

3. Falls alle Ergiebigkeiten genügend klein sind, STOP.

Sonst springe zurück auf 1.

I.9. Algorithmus: (Knotenanalyse hier) (vgl. [16])

In unserem Fall gilt

$$T_{ij} = e^{\lambda_i - x_j - \beta c_{ij}}$$

und damit

$$k_{ij}(T_{ij}) = \lambda_i - x_j = \ln T_{ij} + \beta c_{ij}$$

und man erhält aus den Restriktionen (I.3.2) und (I.3.3) die nachstehenden Gleichungen für λ_i und x_j

$$e^{\lambda_i} = \frac{\hat{O}_i}{\sum_{j=1}^J e^{-x_j} e^{-\beta c_{ij}}} , \quad (I.3.13)$$

$$e^{-x_j} = \frac{\hat{D}_j}{\sum_{i=1}^I e^{\lambda_i} e^{-\beta c_{ij}}} , \quad (I.3.14)$$

die man iterativ im Einzelschrittverfahren löst und die den Gleichungen (I.3.11) und (I.3.12) aus dem Algorithmus I.8 entsprechen.

3.2.1. Kalibrierung im Fall (b)

Im wesentlichen geht man wie im Fall (a) vor, nur daß jetzt der Parametervektor β mit Hilfe eines Abstiegsverfahrens durch Minimierung des Gütekriteriums (I.3.8) ermittelt wird. Für festen Vektor β erhält man die Potentiale aus den Gleichungen (I.3.13) und (I.3.14), wenn man jeweils den Term $e^{-\beta c_{ij}}$ durch $e^{\beta x_{ij}}$ ersetzt.

3.2.3. Kalibrierung im Fall (c)

Im Gegensatz zu den Fällen (a) und (b) braucht man jetzt kein Knotenanalyseverfahren mehr zur Berechnung der Potentiale, da die Restriktionen wegfallen. Mit diesem Fall und den dabei auftretenden Problemen der Eindeutigkeit und Unimodalität beschäftigt sich die vorliegende Arbeit.

Parameter und Potentiale erhält man durch Minimieren des Gütekriteriums

$$\phi(\pi) := \| t(\lambda, \beta) - \hat{t} \| \quad (I.3.15)$$

mit

$$\pi := (\lambda, \beta).$$

Offensichtlich gilt:

$$\min_{\lambda, \beta} \phi(\lambda, \beta) \leq \min_{\beta} \phi(\lambda(\beta), \beta), \quad (\text{I.3.16})$$

d.h. ohne Erfüllung der Restriktionen wird man im allgemeinen eine bessere Approximation der Fahrtenzahlen auf den Pendlerbögen erreichen.

I.10. Bemerkung: (vgl. [31])

Bei guter Approximation des Vektors $\hat{t}$ in irgendeiner Norm werden auch die Knotenbedingungen gut erfüllt, wie wir im folgenden für den Fall der Euklidischen Norm zeigen werden.

Sei also $\phi(\pi) = \|\hat{t}(\pi) - \hat{t}\|_2^2 \quad \forall \pi \in \mathbb{R}^{I+J+M}$.

Dann gilt für eine Komponente π_i von π im Optimum π^* :

$$\frac{\partial \phi}{\partial \pi_i}(\pi^*) = 2 \sum_{i=1}^I \sum_{j=1}^J (T_{ij}(\pi^*) - \hat{T}_{ij}) \frac{\partial T_{ij}}{\partial \pi_i}(\pi^*) = 0. \quad (\text{I.3.17})$$

Im Fall $T_{ij}(\pi) = e^{\lambda_i - x_j} f(\beta, x_{ij})$ gilt dann bei Differentiation nach λ_i und x_j :

$$\sum_{j=1}^J T_{ij}^2(\pi^*) = \sum_{j=1}^J \hat{T}_{ij} T_{ij}(\pi^*) \quad (\text{I.3.18})$$

$$\text{und} \quad \sum_{i=1}^I T_{ij}^2(\pi^*) = \sum_{i=1}^I \hat{T}_{ij} T_{ij}(\pi^*). \quad (\text{I.3.19})$$

Statt der Restriktionen

$$\sum_{j=1}^J T_{ij}(\pi^*) = \sum_{j=1}^J \hat{T}_{ij}$$

$$\text{und} \quad \sum_{i=1}^I T_{ij}(\pi^*) = \sum_{i=1}^I \hat{T}_{ij}$$

werden also die mit $T_{ij}(\pi^*)$ gewichteten Restriktionen (I.3.18)

und (I.3.19) erfüllt. Analog gilt im Minimum statt der Gleichheit der empirischen und Modellgesamtreisekosten

$$\sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij}(\pi^*) = \sum_{i=1}^I \sum_{j=1}^J c_{ij} \hat{T}_{ij}$$

nun die Gleichheit der gewichteten Gesamtreisekosten

$$\sum_{i=1}^I \sum_{j=1}^J c_{ij} T_{ij}^2(\pi^*) = \sum_{i=1}^I \sum_{j=1}^J c_{ij} \hat{T}_{ij} T_{ij}(\pi^*), \quad (I.3.20)$$

wenn man $f(\beta, x_{ij}) := e^{\beta c_{ij}}$ verwendet.

Sei $w_{ij} := T_{ij}(\pi^*) / \hat{T}_{ij}$.

Dann folgt aus (I.3.18) und (I.3.19)

$$\sum_{j=1}^J \frac{1}{w_{ij}} \hat{T}_{ij} = \sum_{j=1}^J \hat{T}_{ij} \quad (I.3.21)$$

$$\sum_{i=1}^I \frac{1}{w_{ij}} \hat{T}_{ij} = \sum_{i=1}^I \hat{T}_{ij} \quad (I.3.22)$$

Daher müssen einige der Gewichtungsfaktoren größer, andere kleiner als 1 sein; es kann nicht der Fall auftreten, daß

$$T_{ij}(\pi^*) > \hat{T}_{ij} \quad \text{oder} \quad T_{ij}(\pi^*) < \hat{T}_{ij}$$

für alle i und j gilt.

§4. Kalibrierung von Cobb-Douglas-Funktionen

Im allgemeinen werden die in der Ökonometrie bei der Kalibrierung von Cobb-Douglas-Funktionen üblichen Gütekriterien aus einer Maximum-Likelihood-Überlegung heraus statistisch begründet. Da sich diese Begründungen teilweise auch auf Gütekriterien in die-

ser Arbeit übertragen lassen, stellen wir in diesem Paragraphen kurz die in der Ökonometrie verwendeten Annahmen über die Verteilung der Meßfehler und daraus resultierende Gütekriterien vor, ohne dabei jedoch auf Einzelheiten (dazu siehe etwa Goldfeld und Quandt [23]) einzugehen, da dies den Rahmen dieser Arbeit sprengen würde.

I.11. Bemerkung:

Gegeben sei ein ökonomisches Modell

$$y = f(A, \beta_1, \dots, \beta_K, x^1, \dots, x^K) = A \prod_{k=1}^K e^{\beta_k x^k} \quad (\text{I.4.1})$$

mit den erklärenden Variablen $x^1, \dots, x^K$ und den unbekannten Parametern $A, \beta_1, \dots, \beta_K$.

I.12. Beispiele:

a) Cobb-Douglas-Nachfrage-Funktionen für Haushalte: [25, S.24]

$$q = f(p, I) = q_0 p^{-\varepsilon} I^{\eta}, \quad (\text{I.4.2})$$

q : nachgefragte Menge;

p : Preis;

I : Einkommen;

q_0, ε, η : Parameter.

b) Cobb-Douglas-Produktions-Funktionen für Firmen: [25, S.25]

$$Y = A K^{\alpha} L^{\beta}, \quad (\text{I.4.3})$$

Y : produzierte Menge;

K : Kapital;

L : Arbeit;

A, α, β : Parameter, $0 < \alpha, \beta < 1, \alpha + \beta = 1$.

I.13. Bemerkung:

Um das allgemeine Modell an eine bestimmte Firma oder einen bestimmten Haushalt anzupassen, kalibriert man die Parameter so, daß die mit ihnen für vorgegebene Werte x_m^k berechneten Modell-ergebnisse y_m die empirischen Werte $\hat{y}_m$ möglichst gut approximieren. Die empirischen Daten sind mit stochastischen Meßfehlern überlagert, über deren Verteilung man, sofern nichts genaueres bekannt ist, üblicherweise eine der folgenden Hypothesen aufstellt:

- die Fehler sind additiv und normalverteilt
- die Fehler sind multiplikativ und log-normalverteilt
- sowohl additive, normalverteilte, als auch multiplikative, log-normalverteilte Fehler treten auf.

Zur Bestimmung der Parameter verwendet man dann häufig die Maximum-Likelihood-Schätzfunktionen.

Wir gehen im folgenden wieder vom Modell (I.4.1) aus und bezeichnen mit

$$y_1, \dots, y_m, \dots, y_M$$

die theoretischen Ergebnisse, die das Modell für vorgegebene Einabewerte x_m^k , $k=1, \dots, K$, $m=1, \dots, M$, liefert, also

$$y_m = A \prod_{k=1}^K e^{\beta_k x_m^k} \quad (I.4.4)$$

I.14. Additiver Fehler: (vgl. [23])

Seien die empirischen Daten von der Form

$$\hat{y} = A \prod_{k=1}^K e^{\beta_k x^k} + v \quad (I.4.5)$$

mit normalverteiltem Fehler v mit Varianz σ_v^2 und Erwartungswert 0.

v hat dann die Dichte

$$g(v) := \frac{1}{\sqrt{2\pi} \sigma_v} \exp\left(-\frac{v^2}{2\sigma_v^2}\right) \quad (\text{I.4.6})$$

und als Erwartungswert von $\hat{y}$ ergibt sich:

$$E(\hat{y}) = A \prod_{k=1}^K e^{\beta_k x^k} \quad (\text{I.4.7})$$

Das Modell ist homoskedastisch, d.h. die Fehlervarianz hängt nicht von den Eingabegrößen x^k ab.

I.15. Multiplikativer Fehler: (vgl. [23])

Seien die empirischen Daten von der Form

$$\hat{y} = A \prod_{k=1}^K e^{\beta_k x^k} e^u =: B e^u \quad (\text{I.4.8})$$

mit normalverteiltem Fehler u (Varianz σ_u^2 , Erwartungswert 0).

Dann ist $B e^u =: z$ log-normalverteilt mit der Dichte

$$f(z) := \frac{1}{\sqrt{2\pi} \sigma_u z} \exp\left(-\frac{(\ln \frac{z}{B})^2}{2\sigma_u^2}\right). \quad (\text{I.4.9})$$

Als Erwartungswert von $\hat{y}$ erhält man

$$E(\hat{y}) = A \prod_{k=1}^K e^{\beta_k x^k} e^{\sigma_u^2/2}, \quad (\text{I.4.10})$$

während

$$\bar{y} := A \prod_{k=1}^K e^{\beta_k x^k} \quad (\text{I.4.11})$$

den Median von $\hat{y}$ darstellt. Das Modell ist heteroskedastisch mit Varianz

$$\text{var}(\hat{y}) = (E(\hat{y}))^2 (e^{\sigma_u^2} - 1). \quad (\text{I.4.12})$$

Die Fehlervarianz hängt also in diesem Fall von den Eingabegrößen x^k ab.

I.16. Gemischter Fehler: (vgl. [23])

Seien die empirischen Werte von der Form

$$\hat{y} = v \cdot e^u + v \quad (\text{I.4.13})$$

mit

$$v = A \prod_{k=1}^K e^{\beta_k x^k}.$$

Die verschiedenen Stichprobenfehler seien unabhängig normalverteilt (u und v wie in (I.4.5) bzw. (I.4.9)). Bei gleichzeitigem Auftreten beider Fehler erhält man als Dichte das Faltungsintegral

$$h(\hat{y}) = \frac{1}{2\pi\sigma_u\sigma_v} \int_0^{\infty} \frac{1}{x} \exp\left(-\frac{1}{2}\left(\frac{\ln \frac{x}{B}}{\sigma_u^2} + \frac{(\hat{y}-x)}{\sigma_v^2}\right)\right) dx, \quad (\text{I.4.14})$$

dessen Wert sich nicht mehr explizit angeben läßt. Deshalb lassen sich auch Erwartungswert und Varianz nicht mehr in einfacher Form ausdrücken.

Bei einem Stichprobenumfang der Länge M für $\hat{y}$ und $\mathbf{x} := (x^1, \dots, x^K)$ erhält man somit als Likelihood-Funktion

$$\mathcal{L} = \prod_{m=1}^M h(\hat{y}_m) \quad (\text{I.4.15})$$

und als Log-Likelihood-Funktion

$$L = \ln \mathcal{L} = \sum_{m=1}^M \ln h(\hat{y}_m), \quad (\text{I.4.16})$$

die man zur Parameterschätzung mittels der Maximum-Likelihood-Methode benötigt

I.17. Kalibrierungsverfahren: (zu I.14-16)

a) additiver Fehler: (I.14) (vgl. [23])

Bekanntlich führt in diesem Fall die Maximum-Likelihood-Methode auf eine Least-Squares-Schätzfunktion, so daß man zur Schätzung

der Parameter A und $\beta_1, \dots, \beta_K$ die Funktion

$$\sum_{m=1}^M \left(\hat{y}_m - A \prod_{k=1}^K e^{\beta_k x_m^k} \right)^2 \quad (I.4.17)$$

minimiert. Dies ist ein nichtlineares Optimierungsproblem.

b) Multiplikativer Fehler: (I.15)

Zur Parameterschätzung logarithmiert man in diesem Fall die Modellgleichung und löst das entstehende **lineare** Problem mit linearer Regression [23]. Man erhält dann jedoch keine unverzerrten Schätzwerte für A ([5][22][33][11]).

c) Gemischter Fehler: (I.16) (vgl. [23])

Man erhält Maximum-Likelihood-Schätzwerte von A und $\beta_1, \dots, \beta_K$ durch Maximieren von L bezüglich dieser Parameter. Da sich das Integral (I.1.14) aber nicht explizit angeben läßt, muß man es M-mal numerisch auswerten, um einen Funktionswert von L zu erhalten. Dies kompliziert die zum Maximieren notwendigen nichtlinearen Optimierungsverfahren und führt zu hohem Rechenaufwand.

Die in Kapitel III dieser Arbeit behandelten Kalibrierungsverfahren mit von der l_2 - Norm des Fehlervektors abgeleiteten Gütekriterien lassen sich damit statistisch durch Maximum-Likelihood-Überlegungen begründen, wenn man Verteilungen des Meßfehlers wie in Abschnitt I.14 und I.15 annimmt.

KAPITEL II:

APPROXIMATION IN DER l_1 -NORM

Dieses Kapitel beschäftigt sich mit der diskreten l_1 -Approximation durch Produkte von Exponentialfunktionen. Dabei untersuchen wir zunächst den Fall nur eines zu kalibrierenden Parameters, der in der Verkehrstheorie zum Beispiel dann von Interesse ist, wenn die Potentiale so bestimmt werden, daß die Konsistenzbedingungen erfüllt sind, und nur ein Parameter β zur Optimierung eines von der l_1 -Norm abgeleiteten Gütekriteriums dient (vgl. Ansatz (I.2.10)). Anschließend beweisen wir für den mehrparametrischen Fall die lineare Quasiunimodalität der sogenannten Einzelfehlerfunktionen.

Aus Gründen einer bequemerem Schreibweise verwenden wir in diesem und den beiden folgenden Kapiteln, die sich nicht direkt auf Verkehrsverteilungsmodelle beziehen, einige andere Bezeichnungen.

II.1. Bezeichnungen:

Gegeben sei eine K -parametrische Familie von approximierenden Funktionen

$$\begin{aligned} y : \mathbb{R}^K \times \mathbb{R}_+^K &\longrightarrow \mathbb{R} \\ (\pi, x) &\longrightarrow y(\pi, x) = \exp\left(\sum_{k=1}^K \pi_k x^k\right), \end{aligned}$$

wobei $x = (x^1, \dots, x^K)^t$ der sogenannte Vektor der Variablen und $\pi = (\pi_1, \dots, \pi_K)^t$ der sogenannte Parametervektor ist.

Von der zu approximierenden Funktion

$$\begin{aligned}\hat{y} : \mathbb{R}^K &\longrightarrow \mathbb{R} \\ \mathbf{x} &\longmapsto \hat{y}(\mathbf{x})\end{aligned}$$

seien nur die M Funktionswerte $\hat{y}_m := \hat{y}(\mathbf{x}_m)$, $m = 1, \dots, M$, bekannt. Die Funktionswerte $y(\pi, \mathbf{x}_m)$ und $\hat{y}(\mathbf{x}_m)$ fassen wir zu den Vektoren $\mathbf{y}$ und $\hat{\mathbf{y}}$ zusammen und lassen die Argumente weg, falls dies nicht zu Mißverständnissen führt. Sei $\hat{y}_m \geq 0$.

II.2. Problem (P1):

Man bestimme die Komponenten des Parametervektors π derart, daß

$$\phi(\pi) := \|\mathbf{y}(\pi) - \hat{\mathbf{y}}\|_1 = \sum_{m=1}^M |y_m(\pi) - \hat{y}_m| \quad (\text{II.C.1})$$

minimal wird.

Im Hinblick auf die numerische Berechnung der Lösung benötigen wir folgende Definitionen:

II.3. Definition: (vgl. [13])

a) Wir nennen $f : I \longrightarrow \mathbb{R}$ mit $[a, b] \subset I \subset \mathbb{R}$ im Intervall $[a, b]$ unimodal, wenn es ein $x^* \in [a, b]$ gibt, so daß f auf $[a, x^*]$ streng antiton und auf $[x^*, b]$ streng isoton ist.

b) Wir nennen $f : D \longrightarrow \mathbb{R}$ mit $G \subset D \subset \mathbb{R}^K$, wobei G abgeschlossen, beschränkt und konvex ist, linear unimodal auf G , wenn f in $\text{conv}(\mathbf{x}_1, \mathbf{x}_2)$ unimodal ist für alle $\mathbf{x}_1, \mathbf{x}_2 \in G$.

c) Wir nennen $f : I \longrightarrow \mathbb{R}$ mit $[a, b] \subset I \subset \mathbb{R}$ im Intervall $[a, b]$ quasiunimodal, wenn es ein Intervall $I^* \subset [a, b]$ gibt, so daß

gilt:

$$f \text{ ist } \begin{cases} \text{streng antiton auf } [a, c] \\ \text{konstant und } \leq \min \{f(c), f(d)\} \text{ auf } I^* \\ \text{streng isoton auf } [d, b] \end{cases}$$

mit $c = \inf I^*$, $d = \sup I^*$.

d) Wir nennen $f : D \longrightarrow \mathbb{R}$ mit $G \subset D \subset \mathbb{R}^K$ linear quasiunimodal auf G , wenn es für beliebige $x_1, x_2 \in G$ quasiunimodal auf $\text{conv}(x_1, x_2)$ ist.

§1. Der einparametrische Fall: $K = 1$

Wie schon in der Einleitung angedeutet, ist es nicht möglich, den einparametrischen Fall der Approximation mit

$$y(\beta, x) := e^{\beta x} \quad (\text{II.1.1})$$

als Spezialfall der diskreten l_1 -Approximation mit Exponentialsummen unterzuordnen. Dies zeigt das folgende Beispiel.

II.4. Beispiel:

Gegeben seien die beiden zu approximierenden Punkte

$$(x_1, \hat{y}_1) = (0.2/0.2) \text{ und } (x_2, \hat{y}_2) = (1.0/1.0).$$

Dann gibt es genau eine l_1 -beste Approximierende der Form

$$L(A, x) := a e^{\beta x} \quad (\text{Exponentialsumme der Länge 1})$$

$$\text{mit } a = (0.2)^{1.25}$$

$$\text{und } \beta = -\ln a.$$

Weitere lokal-beste Approximierende dieser Form gibt es nicht mehr. Es existieren jedoch zwei l_1 -lokal-beste Approximierende der Form (II.1.1), nämlich

$$\text{mit } \beta = 0$$

$$\text{und } \beta = -5 \ln 5 \approx -8.04.$$

Der erste Wert von β liefert dabei die global beste Approximierende.

Ebenso ist es möglich, zwei und (bei Verwendung mehrerer zu

approximierender Punkte) mehr global beste Approximierende zu erhalten, wenn man mit Funktionen des Typs (II.1.1) approximiert.

Anhand graphischer Darstellungen geben wir nun einige Bezeichnungen an, die wir im folgenden verwenden wollen.

II.3. Definition:

Unter der Gesamtfehlerfunktion verstehen wir das Gütekriterium $\phi(\pi)$, unter der m-ten Einzelfehlerfunktion die Funktion

$$\varphi_m(\pi) := |y(\pi, x_m) - \hat{y}_m| = |y_m(\pi) - \hat{y}_m|. \quad (\text{II.1.2})$$

II.4. Bemerkung:

Die Einzelfehlerfunktion $\varphi_m(\pi)$ stellt den Fehler dar, um den die approximierende Funktion $y(\pi, x)$ an der Stelle x_m von dem zu approximierenden Wert abweicht.

Im Fall eines einzigen Parameters bevorzugen wir β statt π und schreiben:

$$\varphi_m(\beta) = |e^{\beta x_m} - \hat{y}_m| \quad (\text{II.1.3})$$

Die folgenden graphischen Darstellungen zeigen das mögliche Aussehen von Einzelfehler-

funktionen, wenn $\hat{y}_m < 1$

ist.

Offensichtlich gilt:

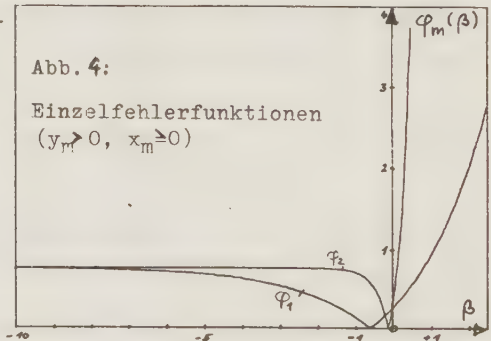
$$\lim_{\beta \rightarrow -\infty} \varphi_m(\beta) = \hat{y}_m$$

$$\lim_{\beta \rightarrow +\infty} \varphi_m(\beta) = +\infty$$

$$\varphi_m\left(\frac{\ln \hat{y}_m}{x_m}\right) = 0.$$

Abb. 4:

Einzelfehlerfunktionen
($y_m > 0, x_m \geq 0$)



Wir bezeichnen

$$\beta_m := \frac{\ln \hat{y}_m}{x_m}$$

als Nullstelle bzw.

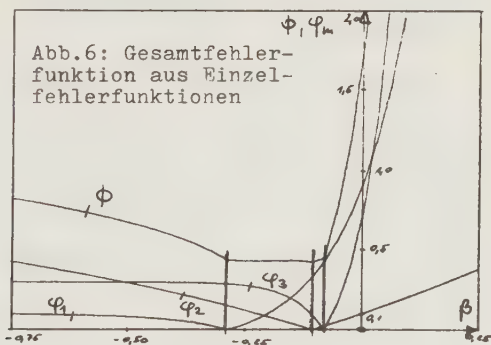
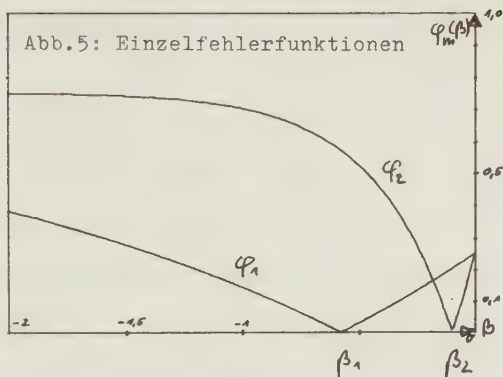
Knickstelle der m-ten
Einzelfehlerfunktion.

Für $\hat{y}_m < 1$ und $x_m > 0$
liegt β_m auf der nega-
tiven Halbachse.

Die Gesamtfehlerfunktion

$$\phi(\beta) = \sum_{m=1}^M \varphi_m(\beta)$$

entsteht durch Überlagerung von M derartigen Einzelfehlerfunk-
tionen und ist im allgemeinen eine Funktion mit M Knickstellen
und M+1 Differenzierbarkeitsintervallen.



II.5. Konvention:

Seien die Paare $(x_m, \hat{y}_m)$ im folgenden stets so indiziert, daß
gilt: $\beta_m \leq \beta_{m+1}$, $1 \leq m \leq M-1$.

II.6. Bemerkung:

Die Minimalstelle der Gesamtfehlerfunktion $\phi(\beta)$ liegt innerhalb des Intervalls $[\beta_1, \beta_M]$, da

$$\begin{aligned}\phi(\beta) &> \phi(\beta_1) \text{ für } \beta \in (-\infty, \beta_1) \text{ und} \\ \phi(\beta) &> \phi(\beta_M) \text{ für } \beta \in (\beta_M, +\infty).\end{aligned}$$

Diese Eigenschaft von ϕ legt folgende Vorgehensweise beim Minimieren von ϕ nahe: man überprüfe, ob im Innern der Intervalle $[\beta_m, \beta_{m+1}]$, $1 \leq m \leq M-1$, relative Minima der Gesamtfehlerfunktion vorliegen. Ist dies der Fall, dann berechne man sie. Kann man jedoch ausschließen, daß innerhalb der Intervalle relative Minima existieren, und dafür werden wir hinreichende Kriterien angeben, dann muß man die Gesamtfehlerfunktion nur noch in den M Knickstellen auswerten, um ihr absolutes Minimum zu finden.

1.1 Untersuchung des Spezialfalles $M=3$ mit äquidistanten x -Werten

Wir gehen auf diesen Spezialfall näher ein, um zu zeigen, daß die Unimodalität der Gesamtfehlerfunktion nicht einmal unter sehr einschränkenden Voraussetzungen an die Werte der zu approximierenden Funktion gesichert ist. Daher werden wir später numerische Verfahren zur Untersuchung auf Unimodalität vorstellen.

Die zu approximierende Funktion sei an den drei Stellen x_1 , x_2 und x_3 durch ihre Werte $\hat{y}_1$, $\hat{y}_2$ und $\hat{y}_3$ gegeben. Die Indizierung entspreche der Konvention II.5., so daß

$$\beta_1 : = \frac{\ln \hat{y}_1}{x_1} \leq \beta_2 : = \frac{\ln \hat{y}_2}{x_2} \leq \beta_3 : = \frac{\ln \hat{y}_3}{x_3} \quad (\text{II.1.4})$$

gilt. Die x -Werte mögen äquidistant liegen, d.h.

$$x_{\max} = x_{\min} + 2\Delta x, \quad (\text{II.1.5})$$

$$x_{\text{mit}} = x_{\min} + \Delta x,$$

wobei $x_{\max} := \max \{x_1, x_2, x_3\}$ und $x_{\min} := \min \{x_1, x_2, x_3\}$.

Ferner gelte $x_i > 0$ und $0 < \hat{y}_i < 1$.

Wir interessieren uns nun dafür, unter welchen Voraussetzungen über die x_i und $\hat{y}_i$ die Gesamtfehlerfunktion $\Phi(\beta)$ unimodal wird und wann in den Differenzierbarkeitsintervallen relative Extrema auftreten.

Dazu genügt es, die Funktion

$$\Phi(\beta) = \sum_{i=1}^3 |e^{\beta x_i} - \hat{y}_i| \quad (\text{II.1.6})$$

in den beiden Intervallen (β_1, β_2) und (β_2, β_3) zu untersuchen.

Differentiation nach β und Ausklammern des Faktors

$$x_{\max} e^{\beta x_{\min}}$$

liefert als notwendige Bedingung für ein relatives Extremum

$$\frac{x_1}{x_{\max}} e^{\beta(x_1 - x_{\min})} + \frac{x_2}{x_{\max}} e^{\beta(x_2 - x_{\min})} - \frac{x_3}{x_{\max}} e^{\beta(x_3 - x_{\min})} = 0, \quad (\text{II.1.7})$$

wobei das obere Vorzeichen im ersten, das untere Vorzeichen im zweiten Intervall gilt. Wir betrachten zunächst den Fall

$x_1 < x_2 < x_3$. Dann gilt für $\lambda := e^{\beta \Delta x}$:

$$\frac{x_{\min}}{x_{\min}+2\Delta x} + \frac{x_{\min}+\Delta x}{x_{\min}+2\Delta x} \lambda - \lambda^2 = 0, \quad (\text{II.1.8})$$

und die Nullstellen λ_1 und λ_2 erfüllen

$$\lambda_1 + \lambda_2 = s \frac{x_{\min}+\Delta x}{x_{\min}+2\Delta x}, \quad s = \mp 1, \quad (\text{II.1.9})$$

$$\lambda_1 \lambda_2 = t \frac{x_{\min}}{x_{\min}+2\Delta x}, \quad t = -1. \quad (\text{II.1.10})$$

Wegen $x_1 > 0$ und $0 < \eta_1 < 1$ liegen mögliche Extremalstellen von ϕ auf der negativen Halbachse, weshalb Nullstellen von (II.1.8) nur dann zulässig sind, wenn gilt: $0 < \lambda_1 < 1$. Man bekommt daraus folgende Auswahlregeln für die Zahl der möglichen Extrema in einem der Intervalle:

$$\left. \begin{array}{l} 0 \text{ Extrema, falls } \lambda_1 < \lambda_2 < 0 \\ \text{oder } \lambda_1 < 0 < 1 < \lambda_2 \\ \text{oder } 1 < \lambda_1 < \lambda_2. \\ 1 \text{ Extremum, falls } \lambda_1 < 0 < \lambda_2 < 1 \\ \text{oder } 0 < \lambda_1 < 1 < \lambda_2. \\ 2 \text{ Extrema, falls } 0 < \lambda_1 < \lambda_2 < 1. \end{array} \right\} \quad (\text{II.1.11})$$

Im Fall einer beliebigen Permutation der Reihenfolge der Größe der x_i ändern nur die Faktoren s und t ihre Vorzeichen. Aus (II.1.9) und (II.1.10) erntnimmt man, daß λ_1 und λ_2 nur vom Quotienten $\Delta x/x_{\min}$ abhängt, so daß Kürzen durch $x_{\min}$ und anschließende Elimination dieses Quotienten den Zusammenhang

$$\lambda_2 = \frac{2t\lambda_1 - st}{s\lambda_1 - 2t} \quad (\text{II.1.12})$$

zwischen λ_1 und λ_2 liefert. Der Fall $\lambda_1 = 2st$, für den der Nenner ver-

schwindet, ist nicht mit den Abschätzungen verträglich, die man aus (II.1.9) und (II.1.10) gewinnt:

$$\frac{1}{2} < |\lambda_1 + \lambda_2| < 1, \quad (\text{II.1.13})$$

$$|\lambda_1 \lambda_2| < 1. \quad (\text{II.1.14})$$

Anwendung der Auswahlkriterien auf die sechs Permutationen der x_i in der Ungleichung $x_1 < x_2 < x_3$ unter Beachtung von (II.1.4) liefert die Tabelle möglicher Konfigurationen der Extrema, die wir auf der nächsten Seite angeben. Die nachfolgenden graphischen Darstellungen zeigen Gesamtfehlerfunktionen aus je drei Einzelfehlerfunktionen.

Zu aporoximieren:

$\frac{x}{y}$	$\frac{1}{0.1}$	$\frac{15}{0.9}$	$\frac{8}{0.3}$
---------------	-----------------	------------------	-----------------

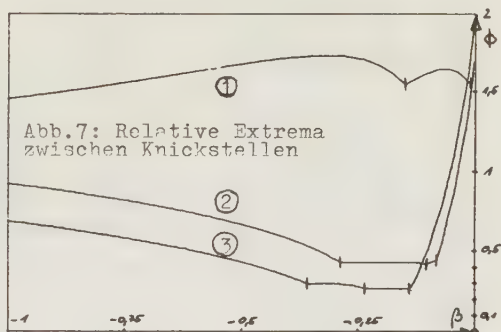
①

$\frac{x}{y}$	$\frac{8}{0.1}$	$\frac{1}{0.9}$	$\frac{15}{0.3}$
---------------	-----------------	-----------------	------------------

②

$\frac{x}{y}$	$\frac{1}{0.7}$	$\frac{15}{0.03}$	$\frac{8}{1/3}$
---------------	-----------------	-------------------	-----------------

③



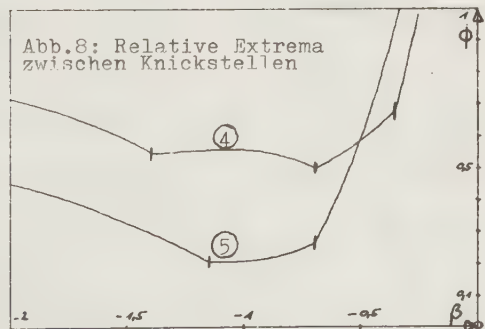
Zu approximieren:

$\frac{x}{y}$	$\frac{1}{1/4}$	$\frac{3}{1/8}$	$\frac{2}{1/2}$
---------------	-----------------	-----------------	-----------------

④

$\frac{x}{y}$	$\frac{3}{0.01}$	$\frac{2}{0.1}$	$\frac{1}{0.5}$
---------------	------------------	-----------------	-----------------

⑤



Ungleichung	$x_1 < x_2 < x_3$	$x_1 < x_3 < x_2$	$x_2 < x_1 < x_3$	$x_2 < x_3 < x_1$	$x_3 < x_1 < x_2$	$x_3 < x_2 < x_1$
Graphische Darstellung						
mögliche Zahl der Extrema	$I_1 \quad I_2$ - - + + - +	$I_1 \quad I_2$ - + + - - -	$I_1 \quad I_2$ + - - + - +	$I_1 \quad I_2$ - + + - + -	$I_1 \quad I_2$ - - - + - -	$I_1 \quad I_2$ - - + + + -
y-Werte	-	$\vartheta_3 < \vartheta_2$	$\vartheta_2 < \vartheta_1$	$\vartheta_2 < \vartheta_1, \vartheta_3 < \vartheta_1$	$\vartheta_3 < \vartheta_1, \vartheta_3 < \vartheta_2$	$\vartheta_3 < \vartheta_2 < \vartheta_1$
Zahlenbsp. 2 Extrema im 1. Intervall	$1 \quad 2 \quad 3$ $.25 \quad .25 \quad .25$	$1 \quad 15 \quad 8$ $.7 \quad .03 \quad .3$	$8 \quad 1 \quad 15$ $.1 \quad .9 \quad .3$	$15 \quad 1 \quad 8$ $.004 \quad .7 \quad .3$	$8 \quad 15 \quad 1$ $.1 \quad .1 \quad .9$	
2 Extrema im 2. Intervall	$1 \quad 2 \quad 3$ $.25 \quad .25 \quad .25$	$1 \quad 3 \quad 2$ $.25 \quad .125 \quad .5$		$45 \quad 7 \quad 26$ $.5 \quad .95 \quad .85$		$45 \quad 26 \quad 7$ $.5 \quad .8 \quad .95$
1 Extremum im 1. Intervall	$7 \quad 26 \quad 45$ $.9 \quad .8 \quad .6$		$26 \quad 7 \quad 45$ $.6 \quad .9 \quad .6$		$2 \quad 3 \quad 1$ $.1 \quad .05 \quad .4$	$3 \quad 2 \quad 1$ $.01 \quad .1 \quad .5$
1 Extremum im 2. Intervall				$6 \quad 4 \quad 5$ $.7 \quad .8 \quad .8$		$6 \quad 5 \quad 4$ $.6 \quad .2 \quad .8$
0 Extrema im 1. Intervall						
0 Extrema im 2. Intervall	$4 \quad 5 \quad 6$ $.8 \quad .8 \quad .8$		$5 \quad 4 \quad 6$ $.7 \quad .8 \quad .8$			

Zur vorstehenden Tabelle:

Bei den graphischen Darstellungen werden auf den Abszissen die x -, auf den Ordinaten die zugehörigen Φ -Werte der zu approximierenden Punkte abgetragen. I_1 bedeutet das Intervall $[\beta_1, \beta_2]$, I_2 das Intervall $[\beta_2, \beta_3]$. In der Zeile mit der möglichen Zahl der Extrema bedeutet "+", daß die links stehende Zahl von Extrema möglich ist, "-", daß sie nicht möglich ist. Die Zahlenschemata sind so zu verstehen:

$$\begin{array}{c|c|c} x_1 & x_2 & x_3 \\ \hline \Phi_1 & \Phi_2 & \Phi_3 \end{array} .$$

Die Auswahlregeln stellen nur notwendige Bedingungen für das Auftreten relativer Extrema dar. Im allgemeinen wird man also auf Grund der Beschränktheit der Intervalle (β_1, β_2) und (β_2, β_3) "seltener" relative Extrema innerhalb der Differenzierbarkeitsintervalle vorfinden, als das nach den Auswahlregeln möglich wäre. Falls jedoch kein λ_1 zwischen 0 und 1 liegt, hat die entsprechende Fehlerfunktion in dem untersuchten Intervall (I_1 bzw. I_2) sicher kein Extremum.

1.2. Unimodalitätskriterien

Da schon in dem soeben betrachteten Spezialfall der Nachweis der Unimodalität bzw. die Untersuchung auf die Anzahl eventuell möglicher relativer Extrema umfangreich war, und da sich die verwendeten Methoden nur unter sehr einschränkenden Bedingungen auf den Fall von mehr als drei Punkten übertragen lassen (z.B. wenn sich die Nullstellen der entstehenden Polynome leicht angeben lassen), beschäftigen wir uns im folgenden mit

numerischen Methoden zur Bestimmung der Anzahl relativer Extrema in den Differenzierbarkeitsintervallen. Dabei geben wir zunächst Kriterien dafür an, daß innerhalb der Differenzierbarkeitsintervalle keine relativen Extrema vorliegen (II.10, II.12, II.18), und beweisen anschließend (Satz II.20), daß dann die nur in Knickstellen der Gesamtfehlerfunktion definierte diskrete Funktion unimodal ist. Diese Eigenschaft vereinfacht die Berechnung des Minimums der Gesamtfehlerfunktion wesentlich.

1.2.1. Sturmsche Ketten

Mit der Methode der Sturmschen Ketten läßt sich die Zahl der reellen Nullstellen eines reellen Polynoms in einem Intervall $[a, b]$ bestimmen.

II.7. Definition:

Sei $f(x)$ ein Polynom und $f_1(x) := f'(x)$. Wir betrachten die Folge von Polynomen $f_i(x)$, deren Glieder ähnlich wie im Euklidischen Algorithmus konstruiert werden:

$$\left. \begin{aligned} f(x) &= g_1(x) f_1(x) - f_2(x) \\ f_1(x) &= g_2(x) f_2(x) - f_3(x) \\ &\vdots \\ f_1(x) &= g_{l+1}(x) f_{l+1}(x) - f_{l+2}(x) \\ f_{l+1}(x) &= g_{l+2}(x) f_{l+2}(x) \end{aligned} \right\} \quad (\text{II.1.16})$$

Dabei ist der Grad von $f_i(x)$ stets größer als der von $f_{i+1}(x)$. $f_{l+2}(x)$ sei der größte gemeinsame Teiler von f und f' .

Die Polynome $f, f_1, f_2, \dots, f_{l+2}$ bilden eine Sturmsche Kette

zu f und f_1 .

II.8. Satz: (Sturm)

Sei f ein reelles Polynom, das in dem Intervall $[a, b]$ nur einfache Nullstellen hat. Sei weiter $f(a) \neq 0$ und $f(b) \neq 0$. Mit $w(x)$ sei die Zahl der Vorzeichenwechsel in der Folge

$$f(x), f_1(x), \dots, f_{1+2}(x)$$

bezeichnet. Dann gilt:

$$w(a) - w(b) = \text{Zahl der Nullstellen von } f \text{ in } [a, b].$$

Beweis: z.B. in [21].

Da die Sturmschen Ketten für Polynome definiert sind, müssen wir die zu minimierende Funktion

$$\phi(\beta) = \sum_{m=1}^M |e^{\beta x_m} - y_m| \quad (\text{II.1.17})$$

umformen. Dazu dient der folgende

II.9. Satz:

Seien die x_m in der Gleichung (II.1.17) von der Form

$$x_m = x_{\min} + \alpha_m \Delta x, \quad (\text{II.1.18})$$

wobei $\alpha_m \in \mathbb{N}$ und Δx der größte gemeinsame Teiler der Zahlen

$x_m - x_{\min}$, $1 \leq m \leq M$, ist.

Dann ist die Zahl der relativen Extrema von

$$\phi(\beta) = \sum_{m=1}^M |e^{\beta x_m} - y_m| \quad (\text{II.1.17})$$

in dem Intervall $[\beta_i, \beta_{i+1}]$, $1 \leq i \leq M-1$, höchstens gleich der Zahl der Nullstellen des Polynoms

$$f_i(\lambda) := \sum_{m=1}^i x_m \lambda^{\alpha_m} - \sum_{m=i+1}^M x_m \lambda^{\alpha_m} \quad (\text{II.1.19})$$

im Intervall $[\lambda_i, \lambda_{i+1}]$ mit $\lambda_i := e^{\beta_i \Delta x}$ und $\lambda := e^{\beta \Delta x}$.

Beweis:

Im Intervall $[\beta_i, \beta_{i+1}]$ hat ϕ die Form

$$\phi_i(\beta) := \sum_{m=1}^i (x_m e^{\beta x_m} - y_m) - \sum_{m=i+1}^M (x_m e^{\beta x_m} - y_m) \quad (\text{II.1.20})$$

und die Ableitung

$$\phi_i'(\beta) := \sum_{m=1}^i x_m e^{\beta x_m} - \sum_{m=i+1}^M x_m e^{\beta x_m}. \quad (\text{II.1.21})$$

Wegen (II.1.15) kann man dies in

$$\phi_i'(\beta) = e^{\beta x_{\min}} \left(\sum_{m=1}^i x_m e^{\beta \Delta x \alpha_m} - \sum_{m=i+1}^M x_m e^{\beta \Delta x \alpha_m} \right) \quad (\text{II.1.22})$$

umformen. Notwendig für ein relatives Extremum von ϕ_i ist

$\phi_i'(\beta) = 0$ und daher

$$\sum_{m=1}^i x_m e^{\beta \Delta x \alpha_m} - \sum_{m=i+1}^M x_m e^{\beta \Delta x \alpha_m} = 0, \quad (\text{II.1.23})$$

woraus mit der Definition $\lambda = e^{\beta \Delta x}$ die Behauptung folgt.
a.e.d.

II.10. Korollar:

Mit den Voraussetzungen und Bezeichnungen von Satz II.9 hat ϕ_i dann kein relatives Extremum in dem betrachteten Intervall, wenn f_i keine Nullstelle hat.

II.11. Bemerkung:

Der Nutzen von Satz II.9 liegt vor allem in diesem Korollar, das festzustellen ermöglicht, wenn keine Extrema in den Differenzierbarkeitsintervallen von ϕ auftreten. In diesem Fall

kann man sich bei der Berechnung des Minimums von ϕ auf die Untersuchung der Knickstellen beschränken.

Falls f_i Nullstellen besitzt und $\max_m \alpha_m$ geradzahlig ist, dann hat ϕ_i mindestens ein Extremum. Um aber feststellen zu können, von welcher Art dieses Extremum ist, muß man seine Lage kennen.

Auf Grund des Satzes von Sturm (II.8) ist es aber naheliegend, wie ein elementarer Algorithmus ablaufen könnte:

Durch Intervallschachtelung in $[\lambda_i, \lambda_{i+1}]$ und jeweilige Anwendung des Satzes von Sturm (II.8) lassen sich näherungsweise alle Nullstellen von f_i und damit alle kritischen Punkte von ϕ_i finden. Man muß dazu nur jeweils in der einmal aufgestellten Sturmschen Kette für f_i die Vorzeichenwechsel an den Intervallgrenzen berechnen, die durch die Intervallschachtelung entstanden sind.

1.2.2 Der Residuensatz

Eine Methode zur Bestimmung der Anzahl der relativen Extrema von ϕ_i , die auch dann anwendbar ist, wenn sich die x_m nicht in der Form (II.1.18) darstellen lassen, liefert die Anwendung des Residuensatzes auf ϕ_i' . Es ergibt sich der folgende

II.12. Satz:

Sei $\bar{\phi}_i'$ die analytische Fortsetzung von ϕ_i' von $\mathbb{R}$ auf $\mathbb{C}$.

Wenn

$$\frac{1}{2\pi i} \int_{\gamma} \frac{\bar{\phi}_i'(z)}{\bar{\phi}_i'(z)} dz = 0, \quad (\text{II.1.24})$$

wobei γ der folgende Weg in $\mathbb{C}$ ist,



dann hat ϕ'_1 im Intervall $[\beta_i, \beta_{i+1}]$ keine reelle Nullstelle.

II.13. Bemerkung:

Der obige Satz liefert nur ein hinreichendes Kriterium dafür, daß die Funktion ϕ kein relatives Extremum hat. Insbesondere können komplexe Nullstellen beliebig nahe an der reellen Achse auftreten, wie man an dem Polynom

$$(z - a - id)(z - a + id) = z^2 - 2az + a^2 + d^2$$

erkennt, so daß auch eine Verkleinerung des Abstandes, den die zur reellen Achse parallel laufenden Wegstücke zu dieser Achse haben, bei der numerischen Auswertung des Integrals (II.1.24) nicht immer aus der Tatsache, daß der Wert des Integrals ungleich Null ist, darauf schließen läßt, daß in dem betrachteten Intervall relative Extrema der Gesamtfehlerfunktion vorliegen.

Falls jedoch der Wert des Integrals eine ungerade Zahl ist, dann liegt wenigstens eine reelle Nullstelle vor.

1.2.3. Abschätzungen

Um festzustellen, ob ϕ im Intervall $[\beta_i, \beta_{i+1}]$ relative Extrema hat, untersuchen wir wieder die Ableitung (II.1.21) auf Nullstellen. Dazu definieren wir

$$\phi'_{i+}(\beta) := \sum_{m=1}^1 x_m e^{\beta x_m} \quad (\text{II.1.25})$$

$$\phi'_{i-}(\beta) := \sum_{m=i+1}^M x_m e^{\beta x_m} \quad (\text{II.1.26})$$

Damit gilt:

$$\phi'_i(\beta) := \phi'_{i+}(\beta) - \phi'_{i-}(\beta) \quad (\text{II.1.27})$$

ϕ'_{i+} und ϕ'_{i-} sind streng isotone Funktionen von β .

II.14.Abschätzung: (0. Ordnung)

Im Intervall $[\beta_i, \beta_{i+1}]$ gilt:

$$\phi'_{i+}(\beta_i) - \phi'_{i-}(\beta_{i+1}) < \phi'_i(\beta) < \phi'_{i+}(\beta_{i+1}) - \phi'_{i-}(\beta_i) \quad (\text{II.1.28})$$

(Die Gültigkeit von (II.1.28) folgt aus der Isotonie von ϕ'_{i+} und ϕ'_{i-} .)

Als Tangenten $T\phi'_{i\pm}$ an $\phi'_{i\pm}$ im Punkt $\bar{\beta} \in [\beta_i, \beta_{i+1}]$ erhält man

$$T\phi'_{i\pm}(\beta) := \phi'_{i\pm}(\bar{\beta}) + \phi'_{i\pm}{}'(\bar{\beta})(\beta - \bar{\beta}), \quad (\text{II.1.29})$$

wobei

$$\phi'_{i+}{}'(\beta) := \sum_{m=i+1}^i x_m^2 e^{\beta x_m}$$

und

$$\phi'_{i-}{}'(\beta) := \sum_{m=i+1}^M x_m^2 e^{\beta x_m}.$$

Als Sekanten $S\phi'_{i\pm}$ an $\phi'_{i\pm}$ über dem Intervall $[\beta_i, \beta_{i+1}]$ ergeben sich

$$S\phi'_{i\pm}(\beta) := \phi'_{i\pm}(\beta_i) + \frac{\phi'_{i\pm}(\beta_{i+1}) - \phi'_{i\pm}(\beta_i)}{\beta_{i+1} - \beta_i} (\beta - \beta_i) \quad (\text{II.1.30})$$

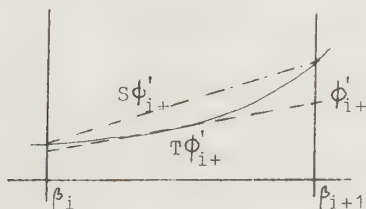


Abb. 9:

II.15. Abschätzung: (1.Ordnung)

Im Intervall $[\beta_i, \beta_{i+1}]$ gilt:

$$T\phi'_{i+}(\beta) - S\phi'_{i-}(\beta) < \phi'_i(\beta) < S\phi'_{i+}(\beta) - T\phi'_{i-}(\beta) \quad (\text{II.1.31})$$

II.16. Bemerkung:

Die Güte der Abschätzung (II.1.31) hängt von der Wahl des Punktes $\bar{\beta}$ ab. Um $\phi'_{i+}(\beta)$ durch Sekante und Tangente auf dem gesamten untersuchten Intervall möglichst eng einschließen zu können, sollte man den Entwicklungspunkt $\bar{\beta}$ so wählen, daß Tangente und Sekante möglichst parallel verlaufen. Außerdem ist $\bar{\beta}$ so zu wählen, daß $T\phi'_{i+}(\beta)$ keine Nullstellen im Intervall $[\beta_i, \beta_{i+1}]$ hat. Die Tangenten $T\phi'_{i+}$ und $T\phi'_{i-}$ können in verschiedenen Punkten an die jeweiligen Kurven angelegt werden.

Als Vorwärtssparabel $V\phi'_{i+}$ bezeichnen wir das zweite Taylorpolynom an ϕ'_{i+} im Punkt β_i , als Rückwärtssparabel $R\phi'_{i+}$ entsprechend das zweite Taylorpolynom an ϕ'_{i+} im Punkt β_{i+1} :

$$V\phi'_{i+}(\beta) := \sum_{m=1}^i x_m (e^{\beta_i x_m} + x_m e^{\beta_i x_m} (\beta - \beta_i)) + \frac{x_m^2}{2} e^{\beta_i x_m} (\beta - \beta_i)^2 \quad (\text{II.1.32})$$

$$R\phi'_{i+}(\beta) := \sum_{m=i+1}^M x_m (e^{\beta_{i+1} x_m} + x_m e^{\beta_{i+1} x_m} (\beta - \beta_{i+1})) + \frac{x_m^2}{2} e^{\beta_{i+1} x_m} (\beta - \beta_{i+1})^2. \quad (\text{II.1.33})$$

Analog sind $V\phi'_{i-}$ und $R\phi'_{i-}$ definiert. Dann gelten

$$\begin{aligned} V\phi'_{i+}(\beta) &\leq \phi'_{i+}(\beta) \leq R\phi'_{i+}(\beta) \\ -R\phi'_{i-}(\beta) &\leq -\phi'_{i-}(\beta) \leq -V\phi'_{i-}(\beta), \end{aligned}$$

so daß wir zusammengefaßt erhalten:

II.17. Abschätzung: (2.Ordnung)

Im Intervall $[\beta_1, \beta_{i+1}]$ gilt:

$$v\phi'_{i+}(\beta) - R\phi'_{i-}(\beta) \leq \phi'_1(\beta) \leq R\phi'_{i+}(\beta) - v\phi'_{i-}(\beta) \quad (\text{II.1.34})$$

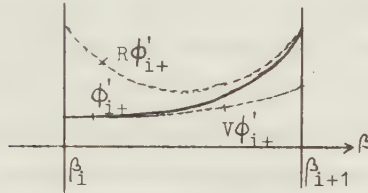


Abb. 10:

II.18. Satz:

Wenn eines der folgenden Kriterien erfüllt ist, dann hat $\phi(\beta)$ im Intervall $[\beta_1, \beta_{i+1}]$ keine relativen Extrema:

$$(1) \quad \phi'_{i+}(\beta_1) - \phi'_{i-}(\beta_{i+1}) > 0 \quad (\text{II.1.35})$$

$$(2) \quad \phi'_{i+}(\beta_{i+1}) - \phi'_{i-}(\beta_1) < 0 \quad (\text{II.1.36})$$

$$(3) \quad T\phi'_{i+}(\beta) - S\phi'_{i-}(\beta) > 0 \quad (\text{II.1.37})$$

$$(4) \quad S\phi'_{i+}(\beta) - T\phi'_{i-}(\beta) < 0 \quad (\text{II.1.38})$$

$$(5) \quad v\phi'_{i+}(\beta) - R\phi'_{i-}(\beta) > 0 \quad (\text{II.1.39})$$

$$(6) \quad R\phi'_{i+}(\beta) - v\phi'_{i-}(\beta) < 0 \quad (\text{II.1.40})$$

Die Beweise folgen unmittelbar aus den Abschätzungen (II.14), (II.15) und (II.16).

II. 19. Bemerkung:

(3) und (4) prüft man nach, indem man die Nullstelle der Geraden berechnet und nachprüft, ob sie im Intervall $[\beta_1, \beta_{i+1}]$

liegt. (5) ist erfüllt, falls es in einem Punkt aus dem Intervall gilt und die Diskriminante der quadratischen Gleichung in β negativ ist, da dann nur komplexe Lösungen auftreten.

Die in Satz II.19 angegebenen Kriterien sind nur hinreichend dafür, daß die Funktion ϕ im Intervall $[\beta_i, \beta_{i+1}]$ keine relativen Extrema hat, d.h. es ist durchaus möglich, daß alle Kriterien verletzt werden und trotzdem keine relativen Extrema auftreten. Mit Hilfe von Taylorpolynomen höherer Ordnung kann man jedoch in analoger Weise Abschätzungen höherer Ordnung gewinnen, um noch genauere Kriterien zu bekommen.

II.20. Satz:

Die Funktion ϕ erfülle in allen Differenzierbarkeitsintervallen $[\beta_i, \beta_{i+1}]$, $1 \leq i \leq M-1$, die Bedingung $\frac{d\phi(\beta)}{d\beta} \neq 0$.

Dann ist die Einschränkung von ϕ auf $\{\beta_i/1 \leq i \leq M\}$ in dem Sinn unimodal, daß ein i_0 existiert mit

$$\phi(\beta_i) > \phi(\beta_{i+1}), \quad i < i_0 - 1,$$

$$\phi(\beta_i) < \phi(\beta_{i+1}), \quad i \geq i_0.$$

Beweis: Wir betrachten die Ableitungen ϕ'_i der Restriktionen von ϕ auf die Intervalle $[\beta_i, \beta_{i+1}]$. Im i -ten bzw. $(i+1)$ -ten Intervall gelten:

$$\phi'_i(\beta) := \sum_{m=1}^i x_m e^{\beta x_m} - \sum_{m=i+1}^M x_m e^{\beta x_m},$$

$$\phi'_{i+1}(\beta) := \sum_{m=1}^{i+1} x_m e^{\beta x_m} - \sum_{m=i+2}^M x_m e^{\beta x_m}.$$

Daraus folgt:

$$\phi'_{i+1}(\beta) = \phi'_i(\beta) + x_{i+1} e^{\beta x_{i+1}}$$

und wegen $x_i > 0$, $1 \leq i \leq M$, gilt weiter

$$\phi'_{i+1}(\beta) > \phi'_i(\beta).$$

Wenn aber dann $\phi'_i(\beta_{i+1}) > 0$ gilt, dann gilt erst recht $\phi'_{i+1}(\beta_{i+1}) > 0$, und daraus folgt mit der Voraussetzung, daß die Ableitung von ϕ im Innern der Intervalle ungleich Null ist, die Behauptung.
q.e.d.

II.21. Bemerkung:

Wenn man mit Hilfe der angegebenen Kriterien festgestellt hat, daß ϕ im Innern der Intervalle **keine** horizontalen Tangenten hat, dann braucht man, um das Minimum von ϕ zu erhalten, nur noch $\phi(\beta_i)$ für $1 \leq i \leq i_0+1$ auszuwerten, und zwar, bis erstmals $\phi(\beta_i) < \phi(\beta_{i+1})$ gilt.

§2. Der mehrparametrische Fall

Die Einzelfehlerfunktionen werden im Fall eines K-dimensionalen Parametervektors π nicht mehr nur in einem Punkt Null, sondern über einer ganzen (K-1)-dimensionalen Hyperebene, nämlich für

$$\sum_{k=1}^K \pi_k x^k = \ln \hat{f}.$$

Das graphische Aussehen zweier solcher Einzelfehlerfunktionen und der aus ihnen entstehenden Gesamtfehlerfunktion zeigen die Abbildungen auf der folgenden Seite.

Im Hinblick auf die Verwendbarkeit von Abstiegsverfahren zur numerischen Berechnung des Minimums interessieren wir uns wieder für die Unimodalität der Gesamtfehlerfunktion.

Abbildung 11:
Einzelfehlerfunktion φ_1

$$\begin{pmatrix} x_{11} \\ x_{12} \\ \hat{y}_1 \end{pmatrix} = \begin{pmatrix} 0.366 \\ 0.275 \\ 3.0 \end{pmatrix}$$

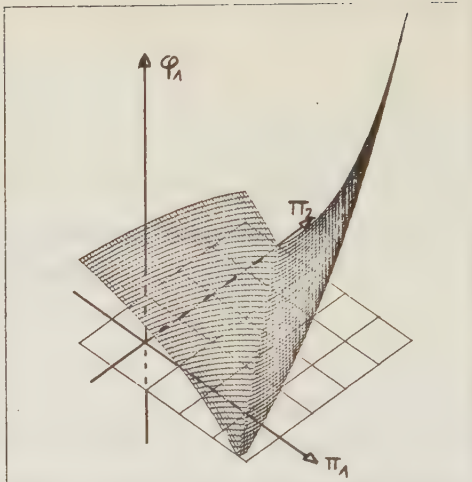


Abbildung 12:
Einzelfehlerfunktion φ_2

$$\begin{pmatrix} x_{21} \\ x_{22} \\ \hat{y}_2 \end{pmatrix} = \begin{pmatrix} 0.069 \\ 0.345 \\ 2.0 \end{pmatrix}$$

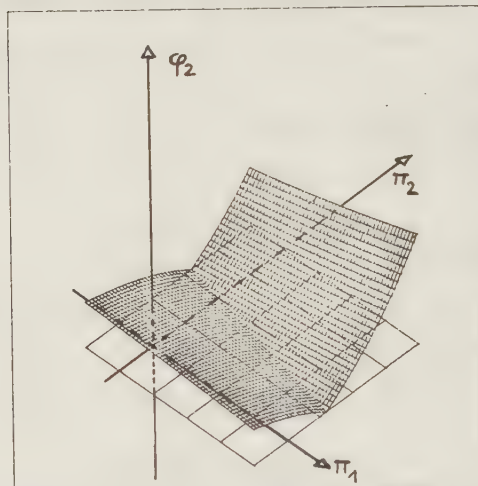
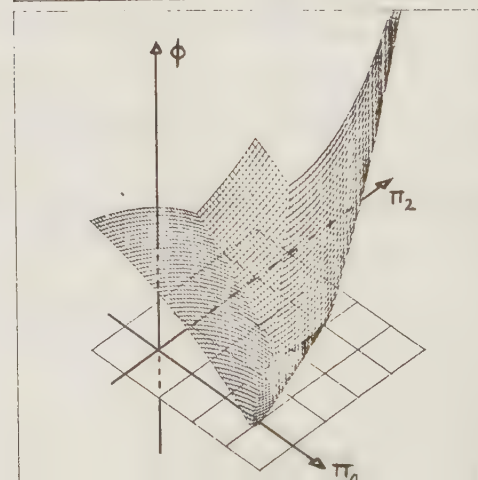


Abbildung 13:
Gesamtfehlerfunktion ϕ
zu den obigen beiden
Einzelfehlerfunktionen

Es gilt jeweils:

$$-1 \leq \pi_1 \leq +1$$

$$-1 \leq \pi_2 \leq +1$$



II.22. Satz:

Die Einzelfehlerfunktionen φ_m sind linear quasiunimodal.

Beweis:

Es gilt
$$\varphi_m(\pi) := \left| \exp \left(\sum_{k=1}^K \pi_k x_m^k \right) - \hat{y}_m \right|. \quad (\text{II.2.1})$$

Wir definieren
$$H_m : \langle \pi, x_m \rangle - \ln \hat{y}_m = 0 \quad (\text{II.2.2})$$

mit
$$\langle \pi, x_m \rangle := \sum_{k=1}^K \pi_k x_m^k. \quad (\text{II.2.3})$$

Diese Hyperebene teilt den Parameterraum $\mathbb{R}^K$ in zwei Halbräume, und es gilt $\varphi_m(\pi) = 0$ für $\pi \in H_m$. O.B.d.A. sei $\ln \hat{y}_m > 0$.

Betrachtet man eine beliebige Gerade

$$\pi(\mu) = \pi' + \mu(\pi'' - \pi') \quad (\text{II.2.4})$$

durch die Punkte π'' und π' des Parameterraumes, so verläuft diese entweder parallel zu H_m , oder sie hat mit H_m einen Schnittpunkt π_s gemeinsam, für den

$$\mu_s := \frac{\ln \hat{y}_m - \langle \pi', x_m \rangle}{\langle \pi'', x_m \rangle - \langle \pi', x_m \rangle} \quad (\text{II.2.5})$$

gilt. Sei $\tilde{\varphi}_m(\mu) := \varphi_m(\pi(\mu))$. Dann folgt:

$$\tilde{\varphi}_m(\mu) = \left| \exp(\pi(\mu), x_m) - \hat{y}_m \right| \quad (\text{II.2.6})$$

und

$$\frac{d\tilde{\varphi}_m(\mu)}{d\mu} = \pm (\langle \pi'', x_m \rangle - \langle \pi', x_m \rangle) \exp(\pi(\mu), x_m), \quad (\text{II.2.7})$$

wobei das obere Vorzeichen für $\mu \in (-\infty, \mu_s)$, das untere für $\mu \in (\mu_s, \infty)$ gilt. (O.B.d.A. sei die Gerade so orientiert, daß $\langle \pi', x_m \rangle < \ln \hat{y}_m < \langle \pi'', x_m \rangle$.) Insgesamt haben wir also

$$\frac{d\tilde{\varphi}_m(\mu)}{d\mu} \begin{cases} < 0, & \text{falls } \mu \in (-\infty, 0) \\ = 0, & \text{falls kein Schnittpunkt ex.} \\ > 0, & \text{falls } \mu \in (0, +\infty). \end{cases} \quad \text{c.q.d.}$$

II.23. Bemerkung:

Zwar sind die Einzelfehlerfunktionen linear quasiunimodale Funktionen, die Gesamtfehlerfunktion muß jedoch nicht linear quasiunimodal sein, wie man sich anhand folgender Beispiele für den Fall $K = 2$ leicht veranschaulicht.

Gegeben seien die sechs Punkte $(x_m^1, x_m^2, y_m)^t$, $m = 1, \dots, 6$:

$(1.0/0.0/0.1)$; $(15.0/0.0/0.9)$; $(8.0/0.0/0.3)$;

$(0.0/1.0/0.1)$; $(0.0/15.0/0.9)$; $(0.0/8.0/0.3)$.

Man bestimme die Parameter π_1 und π_2 derart, daß

$$\sum_{m=1}^6 |e^{\pi_1 x_m^1 + \pi_2 x_m^2 - y_m}| = \min!$$

Die Hyperebenen H_m des Parameterraums, über denen die Einzelfehlerfunktionen zu den ersten (bzw. zweiten) drei Punkten den Wert Null annehmen, verlaufen parallel zur π_2 - (bzw. π_1 -) Achse. Die folgende graphische Darstellung zeigt die Schnittkurven der ersten (bzw. zweiten) drei Einzelfehlerfunktionen und der zugehörigen Summe dieser drei Funktionen mit der (π_1, z) -Ebene (bzw. mit der (π_2, z) -Ebene). Die Gesamtfehlerfunktion zu allen sechs Punkten entsteht durch Addition der beiden Summenfunktionen und hat neun relative Minima, nämlich die Schnittpunkte der Hyperebenen H_1 , H_2 und H_3 mit den Hyperebenen H_4 , H_5 und H_6 .

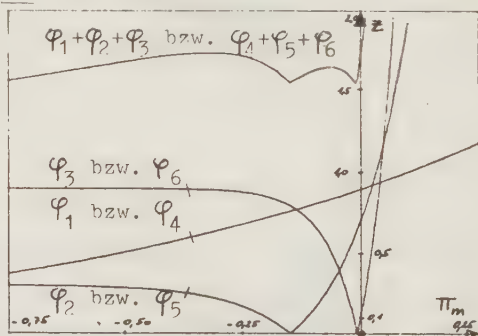


Abbildung 14:
Einzel- und Gesamt-
fehlerfunktionen

Die im vorhergehenden Paragraphen über den einparametrischen Fall angegebenen Kriterien für Unimodalität lassen sich nicht auf den mehrparametrischen Fall verallgemeinern. Bei Residuensatz und Sturmachen Ketten liegt das daran, daß diese Methoden speziell für Funktionen einer Variablen definiert sind und keine mehrdimensionale Entsprechung besitzen. Was die Abschätzungen mit Hilfe der Taylorformel angeht, so lassen sich deren mehrdimensionale Verallgemeinerungen nicht auf das vorliegende Problem anwenden, da es nur in sehr einfachen Fällen möglich ist festzustellen, wie groß die Polyeder sind, die durch die Nullhyperebenen der Einzelfehlerfunktionen festgelegt sind, und damit, in welchem Bereich welche Ableitung von ϕ gilt.

Um daher im mehrparametrischen Fall das globale Minimum zu finden, wird es nötig sein, von verschiedenen Startwerten aus Abstiegsverfahren zu starten, und, falls diese gegen verschiedene relative Minima konvergieren, aus diesen das günstigste auszuwählen.

KAPITEL III:

APPROXIMATION IN DER l_2 -NORM

Dieses Kapitel beschäftigt sich mit der diskreten l_2 -Approximation durch Produkte von Exponentialfunktionen. Wie im vorhergehenden Kapitel untersuchen wir auch hier zunächst den Fall nur eines zu kalibrierenden Parameters und geben Kriterien an, mit denen man numerisch nachprüfen kann, ob die Gesamtfehlerfunktion unimodal ist. Bekanntlich kann man selbst für den Spezialfall der diskreten l_2 -Approximation mit Exponentialsummen der Länge 1, also approximierenden Funktionen der Form

$$L(A, x) := a e^{\beta x} \quad \text{mit } A := (a, \beta), \quad (\text{III.O.1})$$

keine globale Eindeutigkeitsaussage treffen ([8][41]). Dies zeigt auch das folgende Beispiel.

Gegeben seien die fünf zu approximierenden Punkte $(x_m, \hat{y}_m)$:

$$(0/2), (2/0), (2.5/0.535), (3/0), (5/2).$$

Es gibt dazu zwei beste Approximierende der Form (III.O.1),

$$\text{nämlich mit } A_1 = (0.159/0.485)$$

$$\text{und } A_2 = (1.805/-0.485)$$

nach Rundung auf drei Stellen hinter dem Komma.

Andererseits existiert nur eine beste Approximierende der Form

$$\pi(\beta, x) := e^{\beta x} \quad (\text{III.O.2})$$

$$\text{mit } \beta = -0.05,$$

d.h. die l_2 -Approximierende mit "Produkten von Exponentialfunktionen" ist in diesem Spezialfall eindeutig.

Umgekehrt gibt es zu den beiden zu approximierenden Punkten

$$(0.2/0.2) \text{ und } (1/1)$$

genau eine beste l_2 -Approximierende der Form (III.O.1)

mit $a = (1/5)^{5/4}$

und $\beta = -\ln a$.

Neben dieser global besten gibt es keine weiteren lokal-besten l_2 -Approximierenden mehr. Dafür existieren in diesem Fall zwei lokal-beste l_2 -Approximierende der Form (III.0.2)

mit $\beta = -0.035$

und $\beta = -8.01$,

wobei der erste Wert das globale Minimum liefert. Nimmt man als dritten zu approximierenden Punkt noch (0.271/0.271) hinzu, dann bleibt die Approximation mit (III.0.1) eindeutig, mit (III.0.2) dagegen erhält man zwei global beste Approximierende mit demselben Approximationsfehler.

Diese Beispiele zeigen, daß sich im Fall der diskreten l_2 -Approximation keine Eindeutigkeitsaussagen von der Theorie der Approximation mit Exponentialsummen auf die Approximation durch Produkte von Exponentialfunktionen und umgekehrt übertragen lassen.

Es ist aber auch nicht möglich, aus der eindeutigen kontinuierlichen L_2 -Approximierbarkeit einer Funktion durch Approximierende des Typs (III.0.2) auf deren eindeutige diskrete l_2 -Approximierbarkeit zu schließen, wie das folgende Beispiel zeigt.

Die Funktion

$$f(x) := x$$

hat auf dem Intervall (0,1) genau eine beste L_2 -Approximierende der Form

$$g(\beta, x) = e^{\beta x}$$

mit $\beta = -0.805$

(und übrigens auch genau eine der Form (III.0.1) mit $A =$

$= (0.176/0.185)$ (gerundet), aber bei diskreter l_2 -Approximation

in den drei Punkten (0.2/0.2), (0.271/0.271) und (1.0/1.0) zwei global beste Approximierende, nämlich mit $\beta = -5.92$ und $\beta = -0.641$.

In §1 behandeln wir den einparametrischen Spezialfall und geben Kriterien zum numerischen Nachprüfen der Unimodalität an. §2 befaßt sich mit dem mehrparametrischen Fall und stellt die dabei auftretenden Eindeutigkeitsprobleme dar. In §3 formulieren wir die Bestimmung der Potentiale als verallgemeinertes nicht-lineares Eigenwertproblem und lösen dieses mit Fixpunktiteration. Für eine bestimmte gewichtete l_2 -Norm wird dieses Eigenwertproblem linear und seine Lösung eindeutig. §4 schließlich beschäftigt sich mit der näherungsweise Lösung des Problems (P2) mittels exponentieller Regression.

Wir verwenden wieder die Bezeichnungen von Kapitel II und definieren nur die Gesamt- und Einzelfehlerfunktionen neu.

III.1. Problem (P2):

Man bestimme die Komponenten des Parametervektors $\pi \in \mathbb{R}^K$ derart, daß das Gütekriterium

$$\phi(\pi) := \frac{1}{2} \|y(\pi) - \hat{y}\|_2^2 = \frac{1}{2} \sum_{m=1}^M (y_m(\pi) - \hat{y}_m)^2 \quad (\text{III.0.3})$$

minimal wird. Dabei gilt wieder

$$y(\pi) := \exp \left(\sum_{k=1}^K \pi_k x^k \right), \quad (\text{III.0.4})$$

Die Einzelfehlerfunktionen sind von der Form

$$\varphi_m(\pi) := \frac{1}{2} (y_m(\pi) - \hat{y}_m)^2 \quad (\text{III.0.5})$$

§1. Der einparametrische Fall: $K=1$

Zu untersuchen sind die Gesamtfehlerfunktion

$$\phi(\beta) = \frac{1}{2} \|y(\beta) - \hat{y}\|_2^2 = \frac{1}{2} \sum_{m=1}^M (e^{\beta x_m} - \hat{y}_m)^2 \quad (\text{III.1.1})$$

und die Einzelfehlerfunktionen

$$\varphi_m(\beta) = \frac{1}{2} (e^{\beta x_m} - \hat{y}_m)^2. \quad (\text{III.1.2})$$

Abb.15:
Einzelfehlerfunktionen

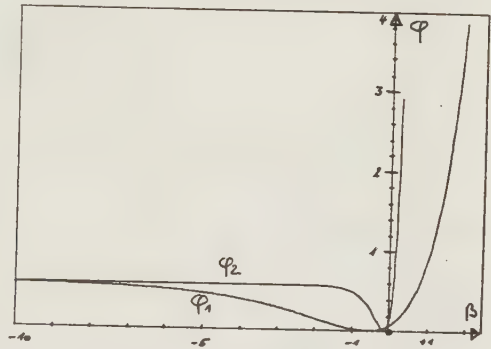
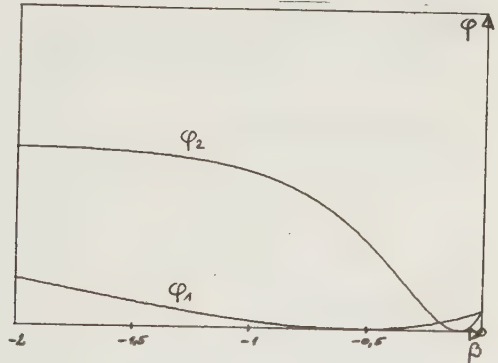


Abb.16:
Einzelfehlerfunktionen



Die Einzelfehlerfunktionen haben jetzt nicht mehr, wie im Fall der l_1 -Norm, "Knicks" an den Nullstellen, sondern sind auf ganz $\mathbb{R}$ stetig differenzierbar. Auch die Gesamtfehlerfunktion ϕ , die durch Addition von M solchen φ_m entsteht, ist auf ganz $\mathbb{R}$ stetig differenzierbar. Um festzustellen, ob ϕ

unimodal ist, brauchen wir also nicht mehrere Teilintervalle einzeln zu untersuchen.

Abb.17:
Gesamtfehlerfunktion
aus Einzelfehlerfunk-
tionen, wobei die Punk-
te

x_m	1	3	2
$\hat{y}_m$	0.25	0.125	0.333

zu approximieren sind.

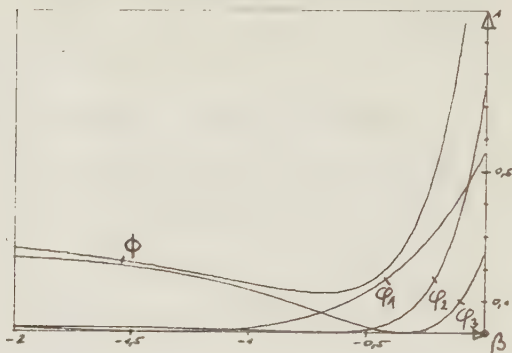


Abb.18:
Verschiedene
Gesamtfehlerfunktionen

x_m	1	15	8
$\hat{y}_m$	0.1	0.9	0.3

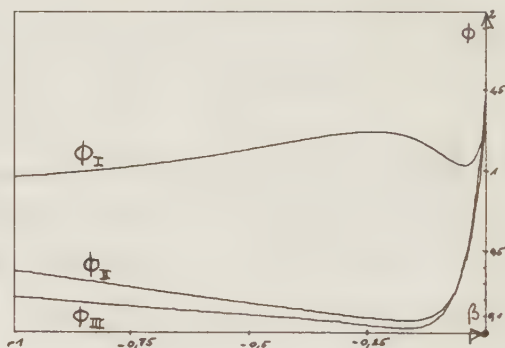
ϕ_I

x_m	8	1	15
$\hat{y}_m$	0.1	0.9	0.3

ϕ_{II}

x_m	1	15	8
$\hat{y}_m$	0.7	0.03	0.333

ϕ_{III}



III.2. Konvention:

Seien die Paare $(x_m, \hat{y}_m)$ im folgenden so indiziert, daß gilt:

$$\beta_m := \frac{\ln \hat{y}_m}{x_m} \leq \beta_{m+1} := \frac{\ln \hat{y}_{m+1}}{x_{m+1}}, \quad 1 \leq m \leq M-1.$$

III.3. Bemerkung:

Die Minimalstelle der Funktion ϕ liegt innerhalb des Intervalls $[\beta_1, \beta_M]$: ϕ ist streng antiton für $\beta \in (-\infty, \beta_1)$ und streng isoton für $\beta \in (\beta_M, \infty)$. Über den Bereich $[\beta_1, \beta_M]$ lassen sich jedoch keine für beliebige Paare $(x_m, \hat{y}_m)$ gültige Monotonie- oder Unimodalitätsaussagen machen: Die Gesamtfehlerfunktion kann

dort relative Extrema haben, wie man aus der nebenstehenden Abb.19 entnimmt.

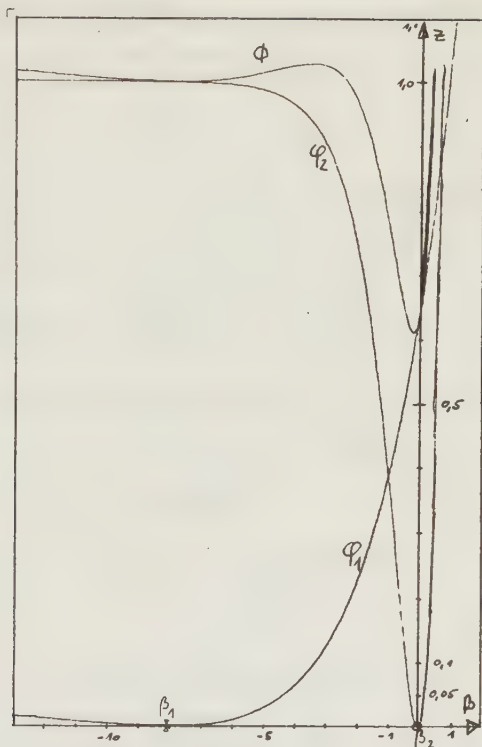


Abb.19:
Gesamtfehlerfunktion
zu den zu approximie-
renden Punkten
(0.2/0.2) und (1/1)

Die Extrema der Gesamtfehlerfunktion werden wieder über die Nullstellen der Ableitung gesucht.

$$\text{Für } \Phi(\beta) = \frac{1}{2} \sum_{m=1}^{M'} (e^{\beta x_m} - \hat{y}_m)^2 \quad \text{ist} \quad (\text{III.1.1})$$

$$\Phi'(\beta) = \frac{d\Phi}{d\beta}(\beta) = \sum_{m=1}^M x_m e^{\beta x_m} (e^{\beta x_m} - \hat{y}_m). \quad (\text{III.1.3})$$

Mit numerischen Methoden wird zunächst die Anzahl der Nullstellen von (III.1.3) bestimmt. Wir verwenden dazu wieder die Sturmschen Ketten und den Residuensatz. Falls mehrere Nullstellen auftreten, kann man sie mit Hilfe dieser beiden

Verfahren genauer lokalisieren und trennen. Anschließend kann man ihre Lage durch ein Abstiegsverfahren eingrenzen.

III.4. Satz:

Seien die x_m in Gleichung (III.1.1) von der Form

$$x_m = x_{\min} + \alpha_m \cdot \Delta x, \quad x_{\min} = \alpha_0 \cdot \Delta x, \quad (\text{III.1.4})$$

wobei $\alpha_m \in \mathbb{N}$ und Δx der größte gemeinsame Teiler der x_m ist.

Dann ist die Zahl der relativen Extrema von (III.1.1) im Intervall $[\beta_1, \beta_M]$ höchstens gleich der Zahl der Nullstellen des Polynoms

$$f(\lambda) = \sum_{m=1}^M x_m (\lambda^{\alpha_0 + 2\alpha_m} - \hat{y}_m \lambda^{\alpha_m}) \quad (\text{III.1.5})$$

im Intervall $[\lambda_1, \lambda_M]$ mit $\lambda := e^{\beta \Delta x}$, $\lambda_i := \hat{y}_i^{\Delta x / x_i}$, $i = 1, M$.

Beweis:

Im Intervall $[\beta_1, \beta_M]$ hat die Ableitung von (III.1.1) die Form

$$\begin{aligned} \phi'(\beta) &= \sum_{m=1}^M x_m (e^{2\beta x_m} - \hat{y}_m e^{\beta x_m}) = \\ &= \lambda^{\alpha_0} \sum_{m=1}^M x_m e^{\beta \Delta x \alpha_0} (e^{\beta \Delta x (\alpha_0 + 2\alpha_m)} - \hat{y}_m e^{\beta \Delta x \alpha_m}) \\ &= \lambda^{\alpha_0} \sum_{m=1}^M x_m (\lambda^{\alpha_0 + 2\alpha_m} - \hat{y}_m \lambda^{\alpha_m}). \end{aligned}$$

q.e.d.

III.5. Korollar:

ϕ hat genau dann ein einziges relatives Extremum, das damit global ist, wenn f nur eine Nullstelle hat.

Beweis:

Da ϕ auf $(-\infty, \beta_1)$ streng antiton und auf $(\beta_M, +\infty)$ streng isoton ist, hat ϕ mindestens ein und, wegen Satz III.4, höch-

stens ein Minimum.

q.e.d.

III.6. Bemerkung:

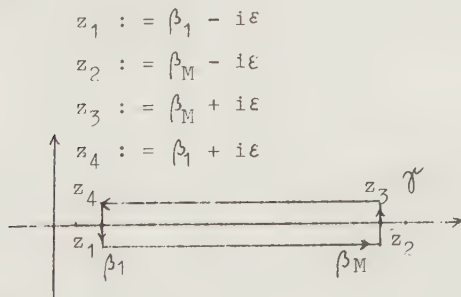
Die Anzahl der Nullstellen des Polynoms f im Intervall $[\lambda_1, \lambda_M]$ erhält man nach dem Satz von Sturm (II.8).

Falls $x_m > 0$, $1 \leq m \leq M$, und $0 < \hat{y}_m < 1$, dann folgt $-\infty < \beta_1 \leq \beta_M < 0$, und es genügt, die Anzahl der Nullstellen von f im Intervall $[0, 1]$ zu bestimmen, statt, wie im Satz III.4, im Intervall $[\beta_1, \beta_M]$.

Wenn mit f_i die Polynome der Sturmschen Kette bezeichnet sind, dann ist $f_i(0)$ gerade die Konstante des Polynoms f_i und $f_i(1)$ die Summe aller seiner Koeffizienten. Man erhält dann schnell die Zahlen $w(0)$ und $w(1)$.

Wenn f mehr als eine Nullstelle hat, dann kann man diese durch fortwährende Intervallhalbierung trennen.

Sei Φ' die analytische Fortsetzung der Funktion ϕ' auf die komplexe Ebene, und es bezeichne γ den in der nachfolgenden Zeichnung angedeuteten geschlossenen Kurvenzug.



Damit ergibt sich unmittelbar der folgende leichte

III.7. Satz:

Wenn

$$\frac{1}{2\pi i} \int_{\gamma} \frac{\bar{\Phi}''(z)}{\bar{\Phi}'(z)} dz = 1, \quad (\text{III.1.6})$$

dann hat ϕ im Intervall $[\beta_1, \beta_N]$ genau ein relatives Extremum. Dieses ist ein globales Minimum.

Beweis:

Unter Berücksichtigung der Bemerkung III.3 liegt im Intervall $[\beta_1, \beta_M]$ mindestens ein relatives Minimum vor, wegen (III.1.6) höchstens eines. q.e.d.

§2. Der mehrparametrische Fall

Wie bei der Approximation in der l_1 -Norm werden auch hier die Einzelfehlerfunktionen über einer ganzen Hyperebene Null, nämlich für

$$\sum_{k=1}^K \pi_k x^k = \ln \hat{y}.$$

Das graphische Aussehen zweier solcher Einzelfehlerfunktionen und der aus ihnen entstandenen Gesamtfehlerfunktion zeigen die Abbildungen auf der folgenden Seite.

Zwar sind die Einzelfehlerfunktionen

$$\varphi_m(\pi) = \frac{1}{2} \left(\exp\left(\sum_{k=1}^K \pi_k x_m^k\right) - \hat{y}_m \right)^2 \quad (\text{III.2.1})$$

linear quasiumodal, die Gesamtfehlerfunktion ist es jedoch im allgemeinen nicht. Als Beispiel für eine nicht linear

Abb.20: l_2 -Norm
Einzelfehlerfunktion φ_1
zu (0.366/0.275/3.0)

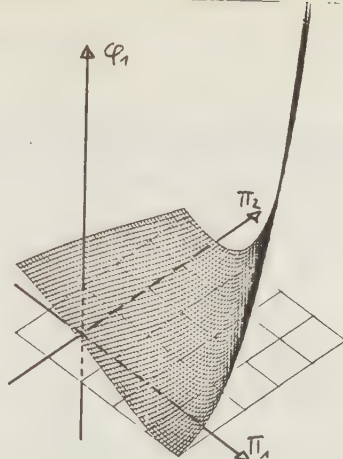


Abb.21: l_2 -Norm
Einzelfehlerfunktion φ_2
zu (0.069/0.345/2.0)

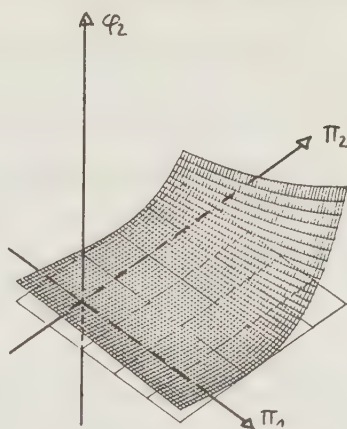
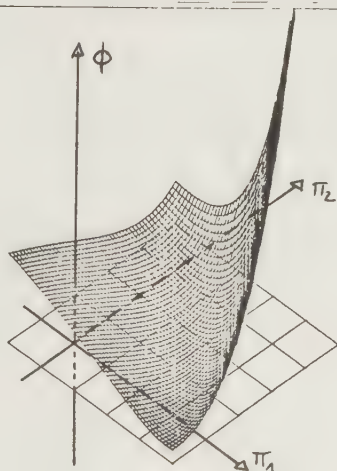


Abb.22: l_2 -Norm
Gesamtfehlerfunktion ϕ
zu den obigen Einzel-
fehlerfunktionen



quasiunimodale Funktion dient die Gesamtfehlerfunktion zu den Punkten $(x_m^1, x_m^2, \hat{y}_m)$, $1 \leq m \leq 4$: (1.0/0.0/1.0); (0.2/0.0/0.2); (0.0/1.0/1.0); (0.0/0.2/0.2):

$$\phi(\pi_1, \pi_2) = \frac{1}{2} \left(\sum_{m=1}^4 (e^{\pi_1 x_m^1 + \pi_2 x_m^2} - \hat{y}_m)^2 \right).$$

Die Nullhyperebenen der Einzelfehlerfunktionen liegen parallel zur π_1 - bzw. π_2 -Achse, die Schnittkurven mit der (π_k, z) -Ebene ($k=1$ oder $k=2$) zeigt die Abb.19 (vergleiche hierzu auch Bemerkung II.23). ϕ hat vier relative Minima und ein relatives Maximum, sowie vier Sattelpunkte. Aus denselben Gründen wie im Fall der l_1 -Norm lassen sich auch hier die im vorigen Paragraphen vorgestellten Kriterien für Unimodalität nicht auf den mehrparametrischen Fall übertragen (vgl. Bem. II.23).

In Ermangelung von a priori-Aussagen über die Unimodalität oder die Zahl auftretender relativer Minima sind wir daher gezwungen, Methoden der globalen Optimierung zu verwenden, um möglichst alle relativen Minima zu berechnen und das günstigste daraus auszuwählen.

Selbst globale Verfahren liefern jedoch nicht mit Sicherheit alle relativen Extrema. Wenn also ein derartiges Verfahren nur ein Extremum findet, so kann man daraus nicht schließen, daß nur eines vorliegt. Ein solches Verfahren ist dennoch eher als die bei Konvexität und Eindeutigkeit üblichen befähigt, viele Extrema zu finden.

Näheres zu den Methoden der globalen Optimierung findet man z.B. bei Dixon und Szegö [12]. Diese beiden Autoren teilen die Verfahren wie folgt ein:

- Techniken der räumlichen Überdeckung (Space Covering Techniques)
- Anziehungsbereichsmethoden
- Trajektorienmethoden
- Verfahren der Funktionenmodifikation
- Zufallsverfahren.

Dixon und Szegö kommen dabei zu dem Schluß, daß im Fall höherer Dimension das einzige praktikable deterministische Verfahren die Trajektorienmethode von Branin ist. Dieser liegt der Gedanke zugrunde, daß man beim Durchlaufen einer Kurve, längs der der Gradient der zu minimierenden Funktion unverändert die gleiche Richtung hat, auch alle Extrema dieser Funktion trifft. Die entsprechende Trajektorie wird durch numerische Integration einer geeigneten Differentialgleichung approximiert.

Die nachfolgenden Überlegungen tragen der speziellen Struktur der Produkte von Exponentialfunktionen, wie sie in der Verkehrstheorie auftreten, Rechnung. Sie lassen sich leichter darstellen, wenn wir das Problem (P2) mit doppelter Indizierung und in folgender Weise formulieren.

III.8. Problem (P2a):

Man bestimme die Komponenten des Parametervektors $\pi \in \mathbb{R}^K$ derart, daß das Gütekriterium

$$\phi(\pi) = \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J (y_{ij}(\pi) - \hat{y}_{ij})^2 \quad (\text{III.2.2})$$

minimal wird. Dabei sei jetzt

$$y_{ij}(\pi) := e^{\lambda_i - x_j} \exp\left(\sum_{m=1}^M \beta_m x_{ij}^m\right), \quad (\text{III.2.3})$$

$$\pi := (\lambda_1, \dots, \lambda_I, x_1, \dots, x_J, \beta_1, \dots, \beta_M)^t \in \mathbb{R}^K,$$

$$\lambda := (\lambda_1, \dots, \lambda_I, x_1, \dots, x_J)^t \in \mathbb{R}^{I+J},$$

$$\beta := (\beta_1, \dots, \beta_M)^t \in \mathbb{R}^M,$$

$$x := (x^1, \dots, x^M)^t \in \mathbb{R}_+^M, \quad \delta_{ij} \text{ Kroneckersymbol},$$

$$x_{ij} := (\delta_{i1}, \dots, \delta_{iI}, -\delta_{1j}, \dots, -\delta_{Jj}, x_{ij}^1, \dots, x_{ij}^M)^t \in \mathbb{R}^K.$$

III.9. Bemerkung:

(P2a) ist ein Spezialfall von (P2) insofern, als der Vektor x_{ij} in

$$y_{ij}(\pi) = \exp(\langle x_{ij}, \pi \rangle) = \exp\left(\sum_{k=1}^K \pi_k x^k\right)$$

von spezieller Form ist. In den ersten I Komponenten tritt genau einmal +1 auf, in den darauffolgenden J genau einmal -1. Nur die letzten M Komponenten können beliebige positive reelle Zahlen sein.

In Abhängigkeit von π ist ϕ dann wieder eine Funktion, deren Unimodalität nicht allgemein gesichert ist. Es können die weiter oben beschriebenen Fälle mit mehreren relativen Extrema auftreten. Die Struktur des Vektors x_{ij} legt es nahe, die Abhängigkeit der Gesamtfehlerfunktion ϕ von den ersten $I + J$ und von den letzten M Komponenten des Parametervektors π getrennt zu untersuchen und zur Bestimmung der beiden Gruppen von Komponenten von π verschiedene, dem Problem angepaßte Verfahren zu verwenden. Unser persönlicher Beitrag bezieht sich dabei auf die erste Klasse von Komponenten, zur Bestimmung der zweiten Klasse verweisen wir auf die Literatur über nichtlineare Optimierungsverfahren.

§3. Fixpunktiteration und verallgemeinerte nichtlineare

Eigenwertmethoden

Wir lösen das Problem (P2a) mittels eines Blockiterationsverfahrens, indem wir abwechselnd bei fest gewähltem λ bezüglich β und bei fest gewähltem β bezüglich λ optimieren. Dieses Verfahren konvergiert stets gegen ein lokales Minimum der Gesamtfehlerfunktion, falls ein solches existiert. Bei der Minimierung der Gesamtfehlerfunktion bezüglich λ tritt somit folgendes Problem auf:

$$\min_{\lambda} \phi(\lambda, \beta)$$

$$\text{mit } \phi(\lambda, \beta) = \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J (e^{\lambda_i - x_j} \exp(\sum_{m=1}^M \beta_m x_{ij}^m) - \hat{y}_{ij})^2. \quad (\text{III.3.1})$$

III.10. Bezeichnungen:

Wir führen folgende Schreibweisen ein:

$$\begin{aligned} l_i &:= e^{\lambda_i}, \quad 1 \leq i \leq I, \\ k_j &:= e^{-x_j}, \quad 1 \leq j \leq J, \\ \mathbf{l} &:= (l_1, \dots, l_I)^t, \\ \mathbf{k} &:= (k_1, \dots, k_J)^t, \\ g_{ij} &:= \exp\left(\sum_{m=1}^M \beta_m x_{ij}^m\right). \end{aligned}$$

Damit wird aus $\phi(\lambda, \beta)$ die Funktion

$$\tilde{\phi}(\mathbf{l}, \mathbf{k}) = \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J (l_i k_j g_{ij} - \hat{y}_{ij})^2. \quad (\text{III.3.2})$$

Notwendig für ein Minimum von (III.3.2) bzgl. $\mathbf{l}$ und $\mathbf{k}$ ist:

$$\nabla \tilde{\Phi}(\mathbf{L}, \mathbf{k}) = \mathbf{0} \in \mathbb{R}^{I+J}. \quad (\text{III.3.3})$$

Gleichbedeutend damit ist das nichtlineare Gleichungssystem

$$\frac{\partial \tilde{\Phi}}{\partial l_i}(\mathbf{L}, \mathbf{k}) = \sum_{j=1}^J k_j g_{ij} (l_i k_j g_{ij} - \hat{y}_{ij}) = 0, \quad 1 \leq i \leq I, \quad (\text{III.3.4})$$

$$\frac{\partial \tilde{\Phi}}{\partial k_j}(\mathbf{L}, \mathbf{k}) = \sum_{i=1}^I l_i g_{ij} (l_i k_j g_{ij} - \hat{y}_{ij}) = 0, \quad 1 \leq j \leq J, \quad (\text{III.3.5})$$

dessen Gleichungen sich nach l_i bzw. k_j auflösen lassen:

$$l_i = \frac{\sum_{j=1}^J \hat{y}_{ij} g_{ij} k_j}{\sum_{j=1}^J g_{ij}^2 k_j^2}, \quad 1 \leq i \leq I, \quad (\text{III.3.6})$$

$$k_j = \frac{\sum_{i=1}^I l_i \hat{y}_{ij} g_{ij}}{\sum_{i=1}^I l_i^2 g_{ij}^2}, \quad 1 \leq j \leq J. \quad (\text{III.3.7})$$

Sei A die Matrix vom Typ (I, J) mit den Elementen

$$a_{ij} := \hat{y}_{ij} g_{ij},$$

der i -ten Zeile $a_{i.}$ und der j -ten Spalte $a_{.j}$. Ferner sei G_i die (J, J) -Gewichtungsmatrix, in deren Diagonale die Elemente $g_{11}, \dots, g_{iJ}$ stehen, und $G_{.j}$ die (I, I) -Gewichtungsmatrix mit $g_{1j}, \dots, g_{Ij}$ in der Diagonale. Alle restlichen Elemente der Gewichtungsmatrizen seien Null. Wir erhalten dann aus (III.3.6) und (III.3.7):

$$l_i = \frac{a_{i.} \cdot \mathbf{k}}{\|G_i \cdot \mathbf{k}\|_2^2}, \quad 1 \leq i \leq I, \quad (\text{III.3.8})$$

$$k_j = \frac{\mathbf{l}^t a_{.j}}{\|G_{.j} \cdot \mathbf{l}\|_2^2}, \quad 1 \leq j \leq J, \quad (\text{III.3.9})$$

so daß wir insgesamt zusammenfassen können:

$$\mathbf{L}(\mathbf{k}) = \mathbf{D}_1(\mathbf{k}) \mathbf{A} \mathbf{k} \quad (\text{III.3.10})$$

mit

$$\mathbf{D}_1(\mathbf{k}) := \begin{pmatrix} 1/\|\mathbf{G}_1\|_2 \|\mathbf{k}\|_2^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1/\|\mathbf{G}_I\|_2 \|\mathbf{k}\|_2^2 \end{pmatrix}$$

und

$$\mathbf{k}(\mathbf{L}) = \mathbf{D}_2(\mathbf{L}) \mathbf{A}^t \mathbf{L} \quad (\text{III.3.11})$$

mit

$$\mathbf{D}_2(\mathbf{L}) := \begin{pmatrix} 1/\|\mathbf{G}_1\|_2 \|\mathbf{L}\|_2^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1/\|\mathbf{G}_J\|_2 \|\mathbf{L}\|_2^2 \end{pmatrix}$$

$$\text{bzw.} \quad \mathbf{k} = \mathbf{D}_2(\mathbf{L}(\mathbf{k})) \mathbf{A}^t \mathbf{D}_1(\mathbf{k}) \mathbf{A} \mathbf{k} \quad (\text{III.3.12})$$

$\mathbf{k}$ ist somit Lösung eines verallgemeinerten nichtlinearen Eigenwertproblems mit dem Eigenwert 1.

Wir betrachten zunächst den speziellen Fall $g_{ij} = 1$, $1 \leq i, j \leq N$.

Man erhält diesen Spezialfall z.B., wenn die zu minimierende Gesamtfehlerfunktion von der Form

$$\tilde{\Phi}(\mathbf{L}, \mathbf{k}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (l_i k_j - g_{ij})^2$$

oder

$$\tilde{\Phi}(\mathbf{L}, \mathbf{k}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (l_i k_j - \frac{\hat{y}_{ij}}{\sigma_{ij}})^2$$

ist. Statt der Gleichungen (III.3.10) und (III.3.11) ergeben sich

$$\mathbf{L} = \frac{\mathbf{A} \mathbf{k}}{\|\mathbf{k}\|_2^2} \quad (\text{III.3.13})$$

und

$$\mathbf{k} = \frac{\mathbf{A}^t \mathbf{L}}{\|\mathbf{L}\|_2^2} \quad (\text{III.3.14})$$

und zusammengefaßt:

$$\|k\| = \frac{A^t A k}{\|k\|_2^2} = \frac{\|k\|_2^4 A^t A k}{\|A^t k\|_2^2 \|k\|_2^2} \quad (\text{III.3.15})$$

bzw. $A^t A k = s k,$ (III.3.16)

wobei $s := \frac{\|A^t k\|_2^2}{\|k\|_2^2}.$

Wir können nun die folgenden Aussagen beweisen:

III.11. Satz:

- (i) Falls $\hat{y}_{ij} > 0, \forall i, j$, dann ist $A^t A$ eine positive Matrix.
- (ii) Der größte Eigenwert einer positiven Matrix ist positiv und einfach, der zugehörige Eigenvektor kann als positiv betrachtet werden.
- (iii) $A^t A$ ist selbstadjungiert (symmetrisch) und hat nur positive reelle Eigenwerte.
- (iv) Zu $A^t A$ existiert eine Orthonormalbasis aus Eigenvektoren.
- (v) Nur der zum größten Eigenwert gehörige Eigenvektor liegt im positiven Orthanten, alle anderen Eigenvektoren haben mindestens eine negative Komponente.

Beweis:

- (i) ist trivial und gilt schon, falls nicht "zu viele" $\hat{y}_{ij}$ Null sind.
- (ii) ist die Aussage des Satzes von Perron [21].
- (iii) und (iv) sind Standardergebnisse der Linearen Algebra.
- (v) folgt aus (i) bis (iv).

o.e.d.

III.12. Korollar:

Die Gesamtfehlerfunktionen der beiden Spezialfälle

$$\tilde{\Phi}(\mathbf{l}, \mathbf{k}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (l_i k_j - \hat{y}_{ij})^2$$

und

$$\tilde{\Phi}(\mathbf{l}, \mathbf{k}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (l_i k_j - \frac{\hat{y}_{ij}}{g_{ij}})^2$$

haben ein eindeutiges Minimum, wenn man einen der unbekannten Parameter l_i oder k_j gleich einer beliebigen positiven reellen Zahl setzt.

Beweis:

Aufgrund der Definition der l_i und k_j ,

$$l_i := e^{\lambda_i} \text{ und } k_j := e^{-x_j},$$

sind nur positive l_i und k_j zulässig. Von allen Lösungen der Eigenvektorgleichung (III.3.16) ist somit nur diejenige brauchbar, die nur aus positiven Komponenten besteht. Dieser Eigenvektor ist jedoch nach Satz III.11(v) bis auf seinen Betrag eindeutig festgelegt. Normiert man eine seiner Komponenten auf 1 (etwa $k_N = 1$), dann liegen auch die anderen Komponenten fest.

q.e.d.

III.13. Bemerkung:

k_N gleich 1 zu setzen bedeutet, x_N gleich 0 zu setzen. In der Terminologie der Netzwerktheorie ist damit der Zielknoten N als Referenzknoten gewählt.

Aus Satz III.11 folgt auch ein Berechnungsverfahren für das Minimum von $\tilde{\Phi}(\mathbf{l}, \mathbf{k})$. Man kann dazu das Verfahren von v. Mises zur Berechnung des größten Eigenwerts einer Matrix verwenden.

III.14. Algorithmus:

(1) Setze $i = 0$. Setze $\mathbf{k}^{(i)} = (1, \dots, 1)^t$.

(2) $\bar{\mathbf{k}}^{(i+1)} := \mathbf{A}^t \mathbf{A} \mathbf{k}^{(i)}, i \in \mathbb{N}_0$.

$$\mathbf{k}^{(i+1)} := \bar{\mathbf{k}}^{(i+1)} / \mathbf{k}_N^{(i+1)}.$$

(3) Wenn der Winkel zwischen zwei aufeinanderfolgenden Vektoren $\mathbf{k}^{(i)}$ und $\mathbf{k}^{(i+1)}$ hinreichend klein ist, setze $\mathbf{k} = \mathbf{k}^{(i+1)}$ und berechne

$$\mathbf{l} := \frac{\mathbf{A} \mathbf{k}}{\|\mathbf{k}\|_2}. \quad (\text{III.3.17})$$

Wenn nicht, ersetze i durch $i+1$ und springe zurück auf (2).

(4) Durch $\lambda_i := \ln l_i$ und $\chi_j := -\ln k_j$ erhält man die gesuchten Potentiale. STOP.

Allgemeiner Fall:

Wir kehren wieder zu dem Problem zurück, $\mathbf{k}$ in der Gleichung

$$\mathbf{k} = \mathbf{D}_2(\mathbf{l}(\mathbf{k})) \mathbf{A}^t \mathbf{D}_1(\mathbf{k}) \mathbf{A} \mathbf{k} \quad (\text{III.3.12})$$

zu bestimmen.

$\mathbf{D}_1$ und $\mathbf{D}_2$ sind Diagonalmatrizen mit streng positiven Diagonalelementen und daher als Quadrate anderer Diagonalmatrizen darstellbar. Wir schreiben kürzer:

$$\begin{aligned} \mathbf{D}_2 \mathbf{A}^t \mathbf{D}_1 \mathbf{A} \mathbf{k} &= \mathbf{k} \\ \mathbf{A}^t \mathbf{D}_1 \mathbf{A} \mathbf{k} &= \mathbf{D}_2^{-1} \mathbf{k} =: \mathbf{R}^2 \mathbf{k} \quad \text{mit } \mathbf{R}^2 = \mathbf{D}_2^{-1} \\ \mathbf{R}^{-1} \mathbf{A}^t \mathbf{D}_1 \mathbf{A} \mathbf{R}^{-1} (\mathbf{R} \mathbf{k}) &= \mathbf{R} \mathbf{k}. \end{aligned}$$

Die Substitution $\mathbf{z} := \mathbf{R} \mathbf{k}$ liefert

$$\mathbf{B} \mathbf{z} = \mathbf{z}$$

mit $\mathbf{B} := \mathbf{R}^{-1} \mathbf{A}^t \mathbf{D}_1 \mathbf{A} \mathbf{R}^{-1}$.

$\mathbf{B}$ ist selbstadjungiert, symmetrisch und positiv, falls $\mathbf{A}$

nur streng positive Elemente enthält.

Allerdings hängt jetzt B , anders als im Spezialfall die Matrix $A^t A$, von z ab. Für einen Lösungsvektor z gelten wieder die Aussagen von Satz III.11: Wegen der Positivität von B ist nach dem Satz von Perron der größte Eigenwert einfach und positiv und der zugehörige Eigenvektor kann als positiv betrachtet werden. Es gibt zu B eine Orthogonalbasis aus Eigenvektoren, weshalb nur ein Eigenvektor, nämlich der zum größten Eigenwert gehörende, im positiven Orthanten liegt.

Da die Elemente von R positive Zahlen sind, bewirkt die Substitution

$$z = R k$$

nur eine "Maßstabsänderung". Die einzelnen Komponenten von k werden mit verschiedenen streng positiven Zahlen multipliziert, ihr Vorzeichen ändert sich also nicht. Wenn daher z als Lösung von

$$B(z) z = z$$

der einzige Eigenvektor im positiven Orthanten ist, dann bleibt nach der Rücksubstitution auch k der einzige Lösungsvektor in R_+^N , d.h. die andern Vektoren der Eigenvektorbasis bleiben nach der Rücksubstitution außerhalb von R_+^N .

Daraus folgt aber nicht die Eindeutigkeit von z , da möglicherweise noch weitere Vektoren z existieren, die im R_+^N liegen und $B(z) z = z$ erfüllen.

Zur numerischen Berechnung einer Lösung der verallgemeinerten Eigenwertprobleme (III.3.12) ist es nicht nötig, den Umweg über die Matrix B zu gehen.

III.15. Algorithmus:

(1) Setze $i = 0$. Setze $\mathbf{k}^{(i)} = (1, \dots, 1)^t$.

(2) $\mathbf{l}^{(i)} := D_1(\mathbf{k}^{(i)}) \mathbf{A} \mathbf{k}^{(i)}$.

(3) $\overline{\mathbf{k}}^{(i+1)} := D_2(\mathbf{l}^{(i)}) \mathbf{A}^t D_1(\mathbf{k}^{(i)}) \mathbf{A} \mathbf{k}^{(i)}$.

$\mathbf{k}^{(i+1)} := \overline{\mathbf{k}}^{(i+1)} / \|\overline{\mathbf{k}}^{(i+1)}\|_1$.

(4) Falls der Winkel zwischen $\mathbf{k}^{(i)}$ und $\mathbf{k}^{(i+1)}$ hinreichend klein ist, dann setze

$$\mathbf{k} := \mathbf{k}^{(i+1)},$$

$$\mathbf{l} := D_1(\mathbf{k}) \mathbf{A} \mathbf{k},$$

falls nicht, ersetze i durch $i+1$ und springe zurück auf (2).

(5) Durch $\lambda_i := \ln(l_i/k_N)$ und $\alpha_j := -\ln(k_j/k_N)$ erhält man die gesuchten Potentiale. STOP.

III.16. Fixpunktsatz: (Brouwer)

Sei $G : \overline{C} \subseteq \mathbb{R}^M \longrightarrow \mathbb{R}^M$ stetig auf der konvexen, kompakten Menge $\overline{C}$ und sei $G\overline{C} \subseteq \overline{C}$. Dann hat G einen Fixpunkt in $\overline{C}$. (Beweis siehe etwa [47, S.161]).

III.17. Korollar:

Die Abbildung

$$G : \overline{C} \subseteq \mathbb{R}^N \longrightarrow \mathbb{R}^N$$

$$\mathbf{k}^{(i)} \longmapsto G \mathbf{k}^{(i)} = \mathbf{k}^{(i+1)}$$

mit $\overline{C} := \{\mathbf{k} \in \mathbb{R}^N / \|\mathbf{k}\|_1 = 1, k_n \geq 0, 1 \leq n \leq N\}$ besitzt mindestens einen Fixpunkt in $\overline{C}$.

Beweis:

$\overline{C}$ ist kompakt und konvex. G ist auf $\overline{C}$ stetig, da D_1 und D_2

stetig auf $\bar{C}$ sind. G ist eine positive Abbildung mit $G\bar{C} \subseteq C \subseteq \bar{C}$. Auch Vektoren auf dem Rand von C werden ins Innere von C abgebildet, Punkte aus dem Innern bleiben im Innern.

Damit erfüllt G die Voraussetzungen von Satz III.16.

a.e.d.

III.18. Bemerkung:

Die Gleichungen (III.3.4) und (III.3.5) lassen sich noch anders deuten, wenn man sie als Funktionen der quadratisch auftretenden Variablen betrachtet. Quadratische Ergänzung liefert dann nämlich

$$\sum_{j=1}^J \left(\frac{k_j}{\frac{1}{g_{ij}}} - \frac{\hat{y}_{ij}}{2l_i} \right)^2 = \left(\frac{1}{4l_i^2} \sum_{j=1}^J \hat{y}_{ij}^2 \right) \quad (\text{III.3.18})$$

und

$$\sum_{i=1}^I \left(\frac{l_i}{\frac{1}{g_{ij}}} - \frac{\hat{y}_{ij}}{2k_j} \right)^2 = \left(\frac{1}{4k_j^2} \sum_{i=1}^I \hat{y}_{ij}^2 \right), \quad (\text{III.3.19})$$

und nach weiterem Umformen:

$$\sum_{j=1}^J \frac{\left(k_j - \frac{\hat{y}_{ij}}{2l_i} \right)^2}{\left(\frac{R_i(l_i)}{g_{ij}} \right)^2} = 1, \quad 1 \leq i \leq I, \quad (\text{III.3.20})$$

$$\sum_{i=1}^I \frac{\left(l_i - \frac{\hat{y}_{ij}}{2g_{ij}k_j} \right)^2}{\left(\frac{S_j(k_j)}{g_{ij}} \right)^2} = 1, \quad 1 \leq j \leq J, \quad (\text{III.3.21})$$

mit den Abkürzungen

$$R_i(l_i) := \frac{1}{2l_i} \sqrt{\sum_{j=1}^J \hat{y}_{ij}^2}, \quad 1 \leq i \leq I,$$

und

$$S_j(k_j) := \frac{1}{2k_j} \sqrt{\sum_{i=1}^I \hat{y}_{ij}^2}, \quad 1 \leq j \leq J.$$

(III.3.20) und (III.3.21) sind Gleichungen von J Ellipsoidschalen im $\mathbb{R}^I$ bzw. von I Ellipsoidschalen im $\mathbb{R}^J$. Für den Spezialfall $g_{ij}=1$ für alle i und j werden diese Ellipsoidschalen zu Sphären, deren Mittelpunkte im positiven Orthanten liegen. Der Koordinatenursprung liegt dabei stets auf allen Ellipsoid- bzw. Kugelschalen.

Wenn die Ortsvektoren der Mittelpunkte von N N -dimensionalen Kugeln linear unabhängig sind und nur (streng) positive Komponenten enthalten und jede der zugehörigen N Sphären $S_1, \dots, S_n, \dots, S_N$ durch den Koordinatenursprung verläuft, dann haben diese Sphären nur noch einen weiteren Punkt gemeinsam. Dies überlegt man sich folgendermaßen. Die "Schnittkurve" von je zwei Sphären definiert eine Hyperebene, die ebenfalls den Nullpunkt enthält und senkrecht zur Verbindungsgerade der entsprechenden Mittelpunkte steht. Da die N Mittelpunktsortsvektoren linear unabhängig sind, sind auch je $N-1$ ihrer Differenzvektoren und damit je $N-1$ Hyperebenen linear unabhängig. Folglich ist der Durchschnitt aller $\frac{N(N-1)}{2}$ derartigen Hyperebenen eine Gerade durch den Koordinatenursprung, die senkrecht zu der durch die N Kugelmittelpunkte aufgespannten Hyperebene verläuft. Sie kann somit nicht tangential zu einer der Kugeln sein, d.h. sie hat mit jeder Kugelschale noch einen weiteren Punkt außer dem Koordinatenursprung gemeinsam.

Das bedeutet aber, daß das durch (III.3.4) und (III.3.5) gegebene Gleichungssystem für den Spezialfall $g_{ij}=1$ (für alle i und j) außer der trivialen nur noch eine weitere Lösung zuläßt und die Funktion $\check{\phi}(l, k)$ aus (III.3.2) nur ein relatives Minimum besitzt.

§4. Lineare l_2 -Approximation logarithmierter Werte

Eine Methode zur Bestimmung der unbekannten Parameter, die aufgrund der Problemstruktur naheliegt und die wegen ihrer Einfachheit häufig Verwendung findet ([25][35]), ist die lineare l_2 -Approximation der logarithmierten Werte durch die logarithmierte Modellgleichung.

III.19. Problem (Pln):

Man bestimme die Komponenten des Parametervektors π derart, daß das Gütekriterium

$$\chi(\pi) := \frac{1}{2} \sum_{m=1}^N (\ln y_m(\pi) - \ln \hat{y}_m)^2 \quad (\text{III.4.1})$$

minimal wird, wobei

$$\ln y_m(\pi) = \sum_{k=1}^K \pi_k x_m^k, \quad 1 \leq m \leq M \quad (\text{III.4.2})$$

gilt. Dieses Minimierungsproblem löst man mit Hilfe der Methode der kleinsten Fehlerquadrate, da es mit der Lösung des überbestimmten linearen Gleichungssystems

$$\sum_{k=1}^K \pi_k x_m^k = \ln \hat{y}_m, \quad 1 \leq m \leq M, \quad (\text{III.4.3})$$

nach den unbekannten Parametern π_k gleichbedeutend ist.

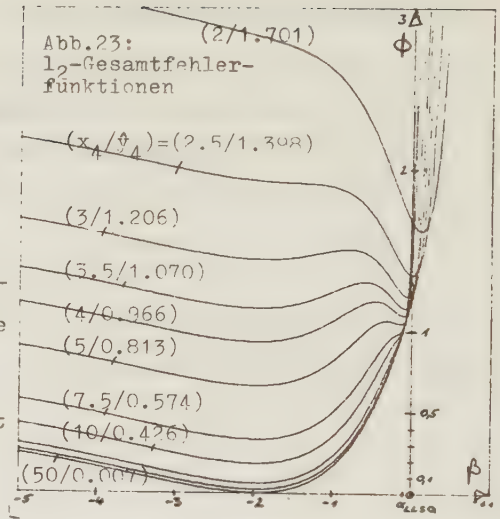
Die auf diese Weise berechneten Parameter ergeben jedoch im allgemeinen kein Minimum des Problems (P2). Sie sind gegen die für (P2) berechneten optimalen Parameter verschoben, wie die folgende Abbildung zeigt. Die Kurven geben die Gesamtfehlerfunktion

$$\phi(\beta) = \sum_{m=1}^M (e^{\beta x_m} - \hat{y}_m)^2$$

im Fall von vier zu approximierenden Werten wieder.

x_m	0.4	0.5	0.6	x_4
$\hat{y}_m$	$e^{-0.8}$	$e^{-1.0}$	$e^{-1.2}$	$\hat{\phi}_4$

Dabei bleiben die ersten drei Wertepaare für jede Kurve gleich. x_4 wird vorgegeben und dann $\hat{\phi}_4$ so bestimmt, daß der mit der Methode der kleinsten Fehlerquadrate berechnete und in der Folge mit α_{LLSQ} bezeichnete Parameter konstant den Wert -0.1 annimmt. Lösen des Problems (Pln) liefert



also stets den Wert α_{LLSQ} für den gesuchten Parameter, während die Lösung des Problems (P2) durch das Minimum der jeweiligen Kurve gegeben wird und offensichtlich teilweise stark von α_{LLSQ} abweicht.

III.20. Bemerkung:

Betrachtet man die $\hat{y}_m$ als Meßwerte einer unbekannten Funktion aus einer mehrparametrischen Funktionenfamilie an den Stellen x_m , dann fallen die Minimalstellen β^* und α_{LLSQ} der Gütekriterien von (P2) und (Pln) zusammen, falls in die $\hat{y}_m$ keine Meßfehler eingehen. Sonst sind sie im allgemeinen gegeneinander verschoben. Diese Verschiebung hängt aber nicht allein von der Größe des Meßfehlers, sondern auch von der Meßstelle x_m ab. Wir veranschaulichen diesen Zusammenhang an einigen graphischen Darstellungen, die jeweils eine Funktion vom Typ

$$x_m^{\pi_1} e^{\pi_2 x_m}$$

mit gegebenem n_1 und n_2 enthalten und dazu die mit dem konstanten Fehler $\pm c$ überlagerten Kurven

$$x_m^{\pi_1} e^{\pi_2 x_m} \pm c,$$

sowie die Kurven der logarithmierten Funktionswerte. Man sieht dabei, wie derselbe Fehler c an verschiedenen Stellen x_m ungleich verzerrt wird.

Abb.24:

$$f(x) = x^2 e^{-x} \pm c$$

$$g(x) = \ln(x^2 e^{-x} \pm c)$$

$$c = \begin{cases} 0.05 \\ 0.0 \\ -0.05 \end{cases}$$

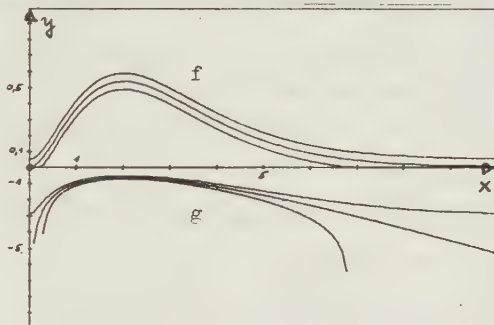


Abb.25:

$$f(x) = x^{-1} e^{-x/2} \pm c$$

$$g(x) = \ln(x^{-1} e^{-x/2} \pm c)$$

$$c = \begin{cases} 0.05 \\ 0.0 \\ -0.05 \end{cases}$$

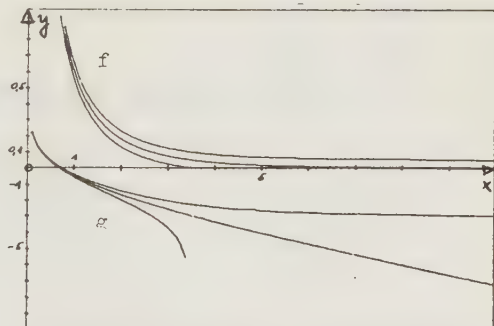
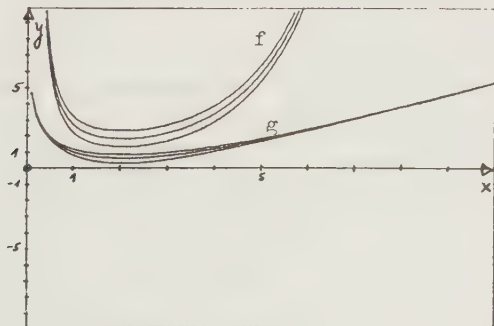


Abb.26:

$$f(x) = x^{-2} e^x \pm c$$

$$g(x) = \ln(x^{-2} e^x \pm c)$$

$$c = \begin{cases} 0.05 \\ 0.0 \\ -0.05 \end{cases}$$



Der obere Maßstab gilt jeweils für die drei oberen, der untere für die drei unteren Kurven.

Besonders deutlich werden die Auswirkungen dieser Verschiebung dann, wenn für kleine x_m die zu approximierenden Funktionswerte $\hat{y}_m$ Meßfehler nach unten und für große x_m Meßfehler nach oben aufweisen. Der für den logarithmierten Fall optimale Parameter α_{LLSQ} ist für den nichtlogarithmierten Fall zu groß, da die Fehler bei den y-Werten für große x-Werte durch das Logarithmieren vergrößert, für kleine x-Werte aber verkleinert werden.

Das Logarithmieren der Werte wirkt sich also ähnlich wie eine Gewichtung der Fehler aus.

III.21. Definition:

Wir führen folgende Bezeichnungen ein, die uns eine bequemere Schreibweise erlauben:

$$y_m := \exp\left(\sum_{k=1}^K \pi_k x_m^k\right) \quad : \text{"theoretischer Wert"}$$

$$\hat{y}_m \quad : \text{"empirischer Wert"}$$

$$c := \hat{y}_m - y_m \quad : \text{"Fehler im nichtlogarithmierten Fall"}$$

$$d := \ln \hat{y}_m - \ln y_m \quad : \text{"Fehler im logarithmierten Fall"}$$

III.22. Satz:

Ein Fehler d im logarithmierten Zustand wird in erster Näherung zu einem Fehler $c \approx y_m d$ im nichtlogarithmierten Fall. Ein Fehler c im exponentiellen Zustand wird in erster Näherung zu einem Fehler $d \approx \frac{1}{y_m} c$ im logarithmierten Zustand.

Beweis:

Das obere Vorzeichen gelte im Fall $\hat{y}_m > y_m$, das untere für $\hat{y}_m < y_m$. Dann folgt:

$$d = \pm \ln \hat{y}_m \mp \ln y_m$$

oder

$$y_m e^{\pm d} = \hat{y}_m = \pm c \pm y_m,$$

so daß

$$c = \mp y_m \pm y_m e^{\pm d}$$

$$= y_m (\pm e^{\pm d} \mp 1)$$

$$= y_m (\mp 1 \pm 1 + d \pm \frac{d^2}{2} + \frac{d^3}{6} \pm \dots)$$

und also

$$c \approx y_m d.$$

n.e.d.

III.23. Bemerkung:

Da ein Fehler c im nichtlogarithmierten Fall in erster Näherung einen Fehler $d = \frac{c}{y_m}$ im logarithmierten Zustand ergibt, wird er dort auch entsprechend bei Anwendung des Fehlerquadratverfahrens berücksichtigt, was einer Gewichtung der m -ten Gleichung mit $\frac{1}{y_m}$ entspricht. Um dem entgegenzuwirken, kann man die m -te Gleichung im logarithmierten Zustand mit y_m gewichten und dann statt des Systems

$$\sum_{k=1}^K \pi_k x_m^k = \ln \hat{y}_m, \quad 1 \leq m \leq M,$$

das System

$$y_m \sum_{k=1}^K \pi_k x_m^k = y_m \ln y_m, \quad 1 \leq m \leq M,$$

bzw. (da y_m nicht bekannt ist, verwendet man stattdessen $\hat{y}_m$)

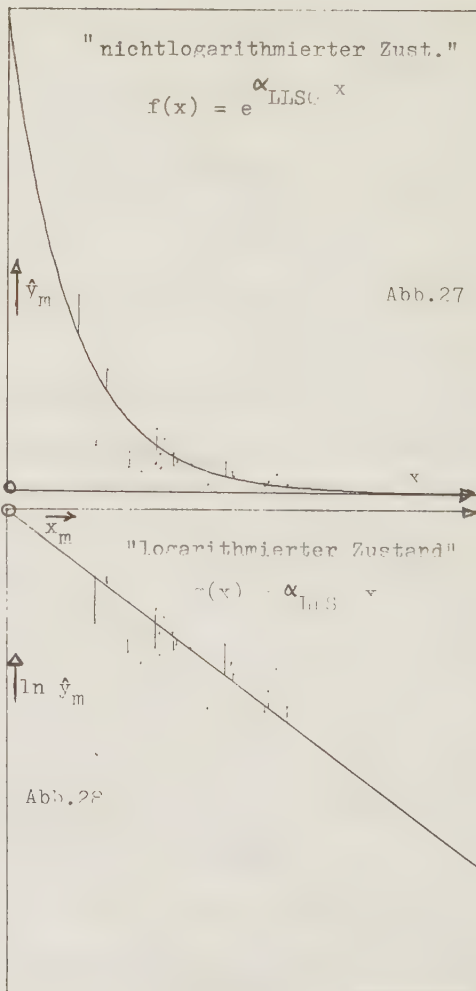
$$\hat{y}_m \sum_{k=1}^K \pi_k x_m^k = \hat{y}_m \ln \hat{y}_m, \quad 1 \leq m \leq M,$$

nach der Methode der kleinsten Fehlerquadrate lösen, um bessere Ausgangswerte für nichtlineare Optimierungsverfahren für

$$\min_{\pi} \frac{1}{2} \sum_{m=1}^M \left(\exp \left(\sum_{k=1}^K \pi_k x_m^k \right) - \hat{y}_m \right)^2$$

zu bekommen.

Die nebenstehenden graphischen Darstellungen zeigen den Effekt, den die Verwendung der Lösung von (Pln) als Näherungslösung von (P2) haben kann. Während im logarithmierten Zustand die l_2 -optimale Nullpunktsgerade so verläuft, daß (III.4.1) minimal wird, ergibt sich für denselben Parameterwert im nichtlogarithmierten Zustand eine Kurve, die für kleine x_m relativ weit nach oben, für große x_m aber nur relativ wenig nach unten von den empirischen Werten abweicht. Die Lösungskurve des Problems (P2) würde in Abb.27 unterhalb von f verlaufen.



KAPITEL IV:

APPROXIMATION IN DER l_∞ -NORM

Dieses Kapitel befaßt sich mit der diskreten Tschebyscheff-Approximation durch Produkte von Exponentialfunktionen. Gerade die Tschebyscheff-Approximation durch Exponentialfunktionen wird in der Literatur ausgiebig behandelt, allerdings vor allem der Fall der diskreten und kontinuierlichen Approximation durch Exponentialsummen

$$L(A, x) := \sum_{i=1}^I a_i e^{\beta_i x}, \quad (IV.0.1)$$

$$A := (a_1, \dots, a_I, \beta_1, \dots, \beta_I)$$

(siehe etwa [3] [32] [28] [29] [30] [38]). Ebenso findet man in der Literatur Existenzaussagen für l_∞ -optimale Approximierende der Form

$$g(a, \pi, x) := a \exp\left(\sum_{k=1}^K \pi_k x^k\right) \quad (IV.0.2)$$

und für Exponentialsummen in mehreren unabhängigen Variablen

$$h(\alpha, \pi, x) := \sum_{j=1}^J p_j(\alpha, x) \exp\left(\sum_{k=1}^K \pi_k x^k\right) \quad (IV.0.3)$$

mit Polynomen p_j ([4] [28] [29] [30]).

Wir beschäftigen uns mit den Ansätzen

$$y(\pi, x) := \exp\left(\sum_{k=1}^K \pi_k x^k\right) \quad (IV.0.4)$$

und

$$\dot{y}(\lambda, \pi, x) := e^{\lambda_1 - x_j} \exp\left(\sum_{k=1}^K \pi_k x^k\right) \quad (IV.0.5)$$

für die approximierenden Funktionen. Beide lassen sich nicht unter einen der Fälle (IV.0.1-3) unterordnen, da in (IV.0.4) der Parameter a fehlt bzw. in (IV.0.5) die Parameter λ_1 und x_j nicht linear einziehen. Die von uns betrachteten Mengen von

Approximierenden bilden im Gegensatz zu den in der Literatur verwendeten Grundmengen keine linearen Räume.

Wir können somit aufgrund der unterschiedlichen Struktur der Approximierenden keine Ergebnisse aus der Literatur übernehmen.

Deshalb beweisen wir in §1 die Konvexität der Lösungsmenge des Problems (P_∞) und die Unimodalität der Fehlerfunktion. In §2 geben wir einen Algorithmus an, der das vorliegende nichtlineare Problem auf eine Folge von linearen Ungleichungssystemen zurückführt und über deren Lösbarkeit mit Hilfe von linearen Programmen entscheidet.

IV.1 Problem (P_∞) :

Man bestimme die Komponenten des Parametervektors $\pi \in \mathbb{R}^K$ derart, daß die Gesamtfehlerfunktion

$$\phi(\pi) := \|y(\pi) - \hat{y}\|_\infty = \max_{1 \leq m \leq M} |y_m(\pi) - \hat{y}_m| \quad (\text{IV.0.6})$$

minimal wird. Dabei gilt wieder

$$y_m(\pi) := \exp \left(\sum_{k=1}^K \pi_k x^k \right). \quad (\text{IV.0.1})$$

Die Einzelfehlerfunktionen sind von der Form

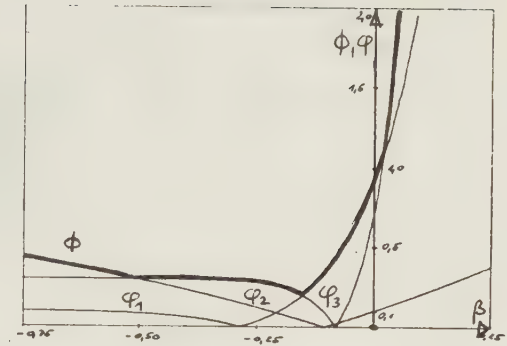
$$\varphi_m(\pi) := |y_m(\pi) - \hat{y}_m| \quad (\text{IV.0.7})$$

und mit denen der l_1 -Norm identisch, die wir in Kapitel II verwendet haben. Für den Zusammenhang zwischen Einzel- und Gesamtfehlerfunktionen gilt:

$$\phi(\pi) = \max_{1 \leq m \leq M} \varphi_m(\pi). \quad (\text{IV.0.8})$$

Die nachfolgende Abbildung gibt für den Fall $K = 1$ die Graphen von Einzel- und Gesamtfehlerfunktionen wieder.

Abb.29:
Einzel- und Gesamt-
fehlerfunktionen bei
diskreter Tscheby-
scheff-Approximation,
 $K = 1, M = 3$.



§1. Konvexität der Minimalmenge

IV.2. Bemerkung:

Wie wir bereits in Kapitel II, Satz II.21 bewiesen haben, sind die Einzelfehlerfunktionen (IV.0.7) linear quasiunimodal. Dabei wird die m -te Einzelfehlerfunktion über der gesamten durch

$$\sum_{k=1}^K \pi_k x_m^k - \ln \hat{y}_m = 0 \quad (\text{IV.1.1})$$

gegebenen Hyperebene Null. Über Hyperebenen, die parallel zu (IV.1.1) verlaufen, z.B.

$$\sum_{k=1}^K \pi_k x_m^k - \ln \hat{y}_m + c = 0, \quad (\text{IV.1.2})$$

ist φ_m konstant.

Auf der folgenden Seite sehen wir graphische Darstellungen von zwei Einzelfehlerfunktionen und der zugehörigen Gesamtfehlerfunktion.

Abb.30:

Tschebyscheff-Einzelfehlerfunktion zu

$$(x_1^1/x_1^2/\hat{y}_1) =$$

$$= (0.366/0.275/3.0):$$

$$\varphi_1(\pi_1, \pi_2) =$$

$$= \left| e^{\pi_1 x_1^1 + \pi_2 x_1^2 - \hat{y}_1} \right|$$

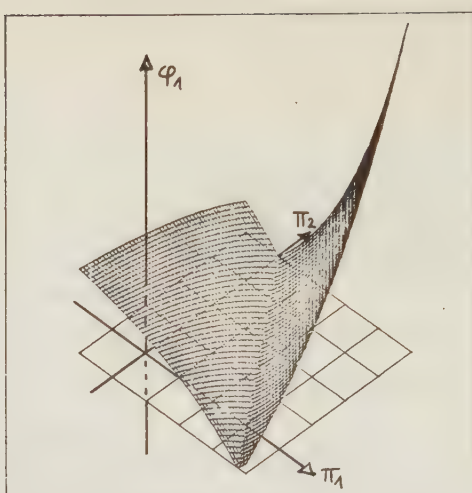


Abb.31:

Tschebyscheff-Einzelfehlerfunktion zu

$$(x_2^1/x_2^2/\hat{y}_2) =$$

$$= (0.069/0.345/2.0):$$

$$\varphi_2(\pi_1, \pi_2) =$$

$$= \left| e^{\pi_1 x_2^1 + \pi_2 x_2^2 - \hat{y}_2} \right|$$

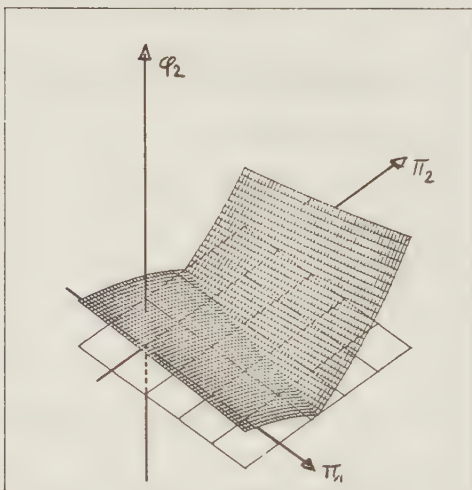
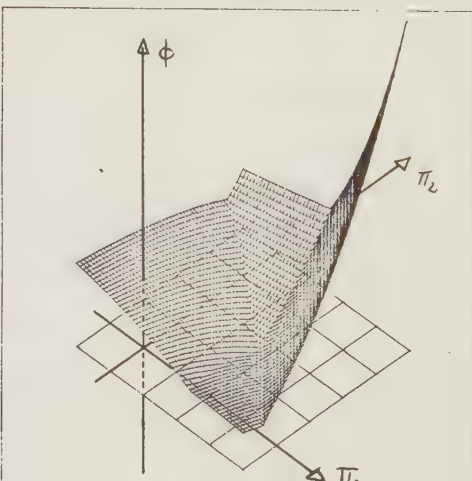


Abb.32:

Tschebyscheff-Gesamtfehlerfunktion zu den obigen Einzelfehlerfunktionen, $M=2$, $K=2$:

$$\phi(\pi_1, \pi_2) = \max \{ \varphi_1, \varphi_2 \}$$



IV.3. Bemerkung:

Bevor wir uns näher mit der Konvexität der Lösungsmenge des Problems (P₀₀) befassen, wollen wir die Menge $N(m, \varepsilon)$, die wir im nachfolgenden Satz IV.4 benötigen und dort allgemein definieren werden, für $K = 2$ veranschaulichen:

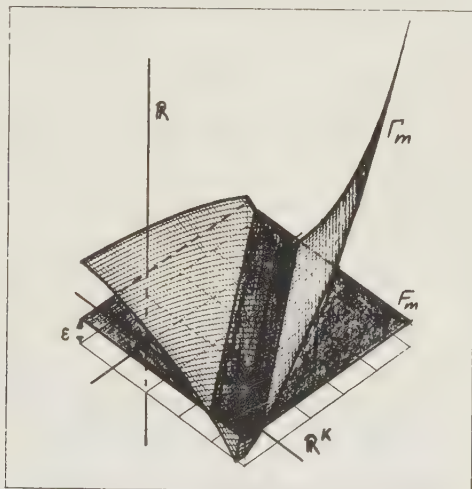


Abb. 33: $N(m, \varepsilon)$

Schneidet man im $\mathbb{R}^{K+1}$ den Graphen Γ_m von φ_m mit der K -dimensionalen, zu $\mathbb{R}^K$ im Abstand ε parallelen Hyperebene F_m und projiziert man die $(K-1)$ -dimensionalen Schnittkurven parallel zur $(K+1)$ ten Achse in den $\mathbb{R}^K$, so erhält man dort die $(K-1)$ -dimensionalen Hyperebenen

$$H_m^+(\varepsilon): \sum_{k=1}^K \pi_k x_m^k = \ln(\hat{\varphi}_m + \varepsilon) \quad (\text{IV.1.3})$$

und

$$H_m^-(\varepsilon): \sum_{k=1}^K \pi_k x_m^k = \ln^*(\hat{\varphi}_m - \varepsilon), \quad (\text{IV.1.4})$$

wobei wir

$$\ln^*(\hat{\varphi}_m - \varepsilon) := \begin{cases} \ln(\hat{\varphi}_m - \varepsilon), & \text{falls } \varepsilon < \hat{\varphi}_m \\ -\infty, & \text{falls } \varepsilon \geq \hat{\varphi}_m \end{cases} \quad (\text{IV.1.5})$$

setzen.

$N(m, \varepsilon)$ ist der Bereich zwischen den Hyperebenen $H_m^+(\varepsilon)$ und $H_m^-(\varepsilon)$. Man erhält für jede Einzelfehlerfunktion φ_m einen Bereich $N(m, \varepsilon)$. Den Durchschnitt dieser Bereiche bezeichnen wir mit $N(\varepsilon)$. Wenn er nicht leer ist, dann ist der Wert jeder Einzelfehlerfunktion φ_m für jeden Vektor im Innern von $N(\varepsilon)$ kleiner gleich ε . Damit ist auch deren Maximum $\phi(\pi)$ kleiner gleich ε .

IV.4. Satz:

Die Lösungsmeße des Problems (P ω) ist ein konvexes Polyeder.

Beweis:

a) Die Menge .

$$N(m, \varepsilon) := \{ \pi \in \mathbb{R}^K / \varphi_m(\pi) \leq \varepsilon \}$$

ist konvex. Denn falls $\varepsilon \geq \hat{y}_m$, dann ist

$$\varphi_m(\pi) = \varepsilon \quad (\text{IV.1.6})$$

die Gleichung einer Hyperebene im $\mathbb{R}^K$, und falls $\varepsilon < \hat{y}_m$, dann ist (IV.1.6) die Gleichung zweier paralleler Hyperebenen. Im ersten Fall ist $N(m, \varepsilon)$ ein Halbraum, im zweiten Fall der Bereich zwischen den Hyperebenen.

b) $N(m, \varepsilon)$ ist eine stetige mengenwertige Abbildung von ε :

$$\begin{aligned} N(m, \varepsilon) &= \{ \pi \in \mathbb{R}^K / \varphi_m(\pi) \leq \varepsilon \} \\ &= \left\{ \pi \in \mathbb{R}^K / \left| \exp\left(\sum_{k=1}^K \pi_k x_m^k\right) - \hat{y}_m \right| \leq \varepsilon \right\} \\ &= \left\{ \pi \in \mathbb{R}^K / \hat{y}_m - \varepsilon \leq \exp\left(\sum_{k=1}^K \pi_k x_m^k\right) \leq \hat{y}_m + \varepsilon \right\} \\ &= \left\{ \pi \in \mathbb{R}^K / \ln(\hat{y}_m - \varepsilon) \leq \sum_{k=1}^K \pi_k x_m^k \leq \ln(\hat{y}_m + \varepsilon) \right\}, \end{aligned}$$

wobei die linke Ungleichung wegfällt, wenn $\hat{y}_m < \varepsilon$.

c) Sei

$$N(\varepsilon) := \bigcap_{m=1}^M N(m, \varepsilon).$$

Dann ist $N(\epsilon)$ entweder leer oder eine konvexe Menge, da es Durchschnitt konvexer Mengen ist.

d) $N(\epsilon)$ ist eine stetige mengenwertige Funktion von ϵ , $\epsilon > 0$. Dies folgt z.B. aus [4, S.119 ff].

e) Wenn $\epsilon' < \epsilon''$, dann folgt aus der Definition von $N(\epsilon)$, daß $N(\epsilon') \not\subseteq N(\epsilon'')$ gilt.

Zusammen mit (d) ergibt sich dann, daß ein $\epsilon_{\min}$ existiert

$$\text{mit} \quad \emptyset \neq N(\epsilon_{\min}) \subseteq N(\epsilon) \quad \forall \epsilon > \epsilon_{\min}$$

$$\text{und} \quad \emptyset = N(\epsilon) \quad \forall \epsilon < \epsilon_{\min}.$$

$N(\epsilon_{\min})$ ist dann das konvexe Polyeder, über dem ϕ minimal wird.

a.e.d.

IV.5. Satz:

Wenn unter der M Vektoren $x_1, \dots, x_M$ von erklärenden Variablen keine K Vektoren linear unabhängig sind, dann ist die Menge $N(\epsilon_{\min})$ einpunktig und somit die Minimalstelle von ϕ eindeutig.

Beweis:

Wir bezeichnen die Menge $\{(\pi, \varphi_m(\pi)) / \pi \in \mathbb{R}^K\}$ als "Einzelfehlerfläche" zu der Einzelfehlerfunktion φ_m . Die Einzelfehlerflächen sind K -dimensionale (nichtlineare) Mannigfaltigkeiten des $\mathbb{R}^{K+1}$.

a) Aus der linearen Quasiumodalität der Einzelfehlerfunktionen φ_m (vgl. Satz II.21) und der Definition von ϕ als Maximum der φ_m folgt, daß das Minimum der "Gesamtfehlerfläche" $\{\pi\phi(\pi) / \pi \in \mathbb{R}^K\}$ auf der Schnittfläche (von zwei oder mehreren) Einzelfehlerflächen angenommen wird.

b) Zwei Einzelfehlerflächen schneiden sich (falls sie nicht zusammenfallen) in (ein oder zwei) $(K-1)$ -dimensionalen Mannigfaltigkeiten. Projektion dieser Mannigfaltigkeiten in den $\mathbb{R}^K$ parallel

zur $(K+1)$ -ten Koordinatenachse liefert eine $(K-1)$ -dimensionale Fläche im $\mathbb{R}^K$.

c) Wegen der linearen Unabhängigkeit der Einzelfehlerfunktionen φ_m für verschiedene Vektoren $\mathbf{x}_m$ von erklärenden Variablen und wegen Bemerkung IV.2 folgt, daß die Projektion der Schnittfläche **zweier** Einzelfehlerflächen nur dann eine $(K-1)$ -dimensionale Ebene im $\mathbb{R}^K$ ist, wenn die zugehörigen Vektoren von erklärenden Variablen linear abhängig sind. Im allgemeinen wird die Projektion der $(K-1)$ -dimensionalen Schnittfläche nur $(K-2)$ -dimensionale Ebenen enthalten.

Ebenso enthält die Projektion der Schnittfläche von K Einzelfehlerflächen nur dann ein Geradenstück, wenn die Schnittfläche selbst schon ein Geradenstück enthält. Nach Bemerkung IV.2 ist dies gleichbedeutend damit, daß die **zugehörigen** Vektoren von Einzelfehlerfunktionen linear abhängig sind. Dies ist ein Widerspruch zur Voraussetzung.

Folglich kann die Menge $N(\varepsilon)$ keine Geradenstücke enthalten. Wegen ihrer Konvexität muß sie daher eipunktig sein. q.e.d.

§2. Numerische Näherungsverfahren zur Lösung

Zur Lösung des Problems (P_∞) geben wir ein Verfahren an, das auf Satz IV.4 beruht und den minimalen Wert $\varepsilon_{\min}$ von ε bestimmt, für den die Menge $N(\varepsilon)$ nicht leer ist.

2.1. Niveaubenenverfahren zur Lösung

Wie im Beweis von Satz IV.4 definieren wir:

$$N(m, \varepsilon) := \left\{ \mathbf{n} \in \mathbb{R}^K / \ln^*(\hat{\varphi}_m - \varepsilon) = \sum_{k=1}^K \pi_k x_m^k = \ln(\hat{\varphi}_m + \varepsilon) \right\},$$

(IV.2.1)

wobei $\ln^*$ und $N(\mathcal{E})$ durch (IV.1.4) bzw. (IV.1.5) gegeben sind.
Zunächst sehen wir zwei äquivalente Formulierungen des Problems $(P\omega)$ an, aus denen wir später das Lösungsverfahren entwickeln.

IV.6. Satz:

Die beiden folgenden nichtlinearen Optimierungsprobleme sind äquivalent zu $(P\omega)$ in dem Sinn, daß sie dasselbe $\mathcal{E}_{\min}$ und dieselbe Minimalmenge liefern:

$(P_{\text{exp}}): \quad \text{Min } \mathcal{E},$

so daß folgendes Ungleichungssystem Lösungen π besitzt:

$$\left| \exp\left(\sum_{k=1}^K \pi_k x_m^k\right) - \hat{y}_m \right| \leq \mathcal{E}, \quad 1 \leq m \leq M. \quad (\text{IV.2.2})$$

$(P_{\text{log}}): \quad \text{Min } \mathcal{E},$

so daß folgendes Ungleichungssystem Lösungen π besitzt:

$$\sum_{k=1}^K \pi_k x_m^k \leq \ln(\hat{y}_m + \mathcal{E}) \quad (\text{IV.2.3})$$

$$\sum_{k=1}^K \pi_k x_m^k \geq \ln^*(\hat{y}_m - \mathcal{E}), \quad (\text{IV.2.4})$$

wobei jeweils $1 \leq m \leq M$ gilt.

Beweis:

Offensichtlich sind (P_{exp}) und (P_{log}) äquivalent, da die Ungleichungssysteme durch Logarithmieren bzw. Exponieren ineinander übergehen. (P_{exp}) ist zu $(P\omega)$ (vgl. IV.0.6) äquivalent.

IV.7. Bemerkung:

Wir sprechen auch bei den äquivalenten Problemen (P_{exp}) und (P_{log}) von nichtlinearen Ungleichungssystemen, da im ersten Fall $\pi \in \mathbb{R}^K$

und im zweiten Fall ϵ nichtlinear eingehen.

Zur Lösung des Problems (Plog) geben wir ϵ fest vor und prüfen die Verträglichkeit der Ungleichungssysteme (IV.2.3) und (IV.2.4) nach, wobei wir uns auf das starke Dualitätstheorem stützen.

Wenn die Ungleichungssysteme (IV.2.3) und (IV.2.4) ein nichtleeres konvexes Polyeder E definieren, dann nimmt eine lineare Kostenfunktion ihr Minimum in einem Extrempunkt von E an.

Der Abstand d der Hyperebenen $H_m^+(\epsilon)$ und $H_m^-(\epsilon)$ hängt streng isoton von ϵ ab, nämlich

$$d := \ln(\hat{y}_m + \epsilon) - \ln^*(\hat{y}_m - \epsilon).$$

Verkleinert man das Niveau ϵ , dann rücken die Hyperebenen näher zusammen.

Sei $\text{extr } N(\epsilon)$ die Menge der Extrempunkte von $N(\epsilon)$. Dann gilt:

$$\pi \in \text{extr } N(\epsilon_0) \implies \pi \notin N(\epsilon) \quad \forall \epsilon < \epsilon_0.$$

Will man daher nur die Verträglichkeit der Ungleichungsrestriktionen von (Plog) nachprüfen, dann ist es günstiger, nicht einen Extrempunkt zu bestimmen, da dieser beim Verkleinern von ϵ keine Lösung mehr darstellt, sondern einen Punkt im Innern von $N(\epsilon)$ zu bestimmen, der noch eine Lösung des Ungleichungssystems bleibt, solange ϵ nicht zu klein wird.

Außerdem kann man mit Hilfe eines geeignet gewählten Punktes π im Innern von $N(\epsilon)$ abschätzen, wie weit man ϵ mindestens noch verkleinern kann, bis π auf dem Rand des Polyeders zu liegen kommt.

Wir geben nun zunächst ein Verfahren an, mit dem man eine Lösung von (Plog) im Innern von $N(\epsilon)$ erhält, und anschließend

eine von dieser Lösung abhängige Abschätzung dafür, wie weit man ε mindestens noch verkleinern kann, so daß die Ungleichungsrestriktionen verträglich bleiben.

IV.9. Hilfsproblem ($P\mathcal{E}_n$) für festes $\mathcal{E}_n$:

Max ζ

unter den Nebenbedingungen

$$\sum_{k=1}^K \pi_k x_m^k + \zeta \leq \ln (\hat{y}_m + \mathcal{E}_n), \quad (\text{IV.2.9})$$

$$\sum_{k=1}^K \pi_k x_m^k - \zeta \geq \ln (\hat{y}_m - \mathcal{E}_n), \quad (\text{IV.2.10})$$

wobei $1 \leq m \leq M$.

IV.10. Bemerkung: (vgl. [43])

Die zusätzliche Variable ζ sorgt dafür, daß eine Lösung π des Systems der Ungleichungsrestriktionen, falls sie überhaupt existiert, von allen Hyperebenen, die $N(\mathcal{E}_n)$ definieren, mindestens einen Abstand ζ hat und damit, falls $\zeta > 0$, im Innern von $N(\mathcal{E}_n)$ liegt.

$\zeta = 0$ bedeutet, daß sich ε nicht mehr weiter verkleinern läßt. Der Abstand einer Lösung π des Hilfsproblems zu einer der $N(\mathcal{E}_n)$ definierenden Hyperebenen ist bereits Null, π liegt auf dieser Hyperebene.

$\zeta < 0$ bedeutet, daß der Punkt π in Bezug auf mindestens eine Hyperebene "auf der falschen Seite" liegt. Das Ungleichungssystem ist inkonsistent.

Wir veranschaulichen diese drei Fälle für $K = 1$.

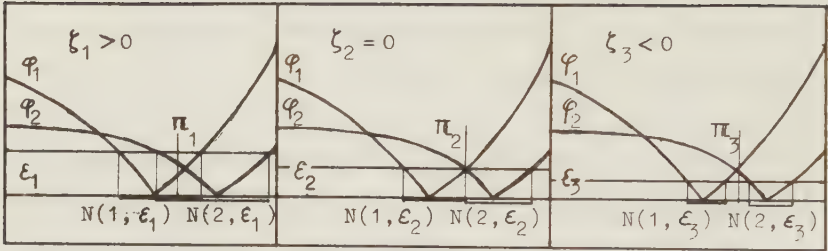


Abb.34

IV.11. Satz:

Sei π_0^* eine Lösung des Problems (P_{ε_0}) und ζ_0^* der zugehörige Minimalabstand zu einer der Hyperebenen (IV.2.7) oder (IV.2.8).

Wenn $\zeta_0^* > 0$, dann kann man ε_0 noch mindestens auf $\varepsilon_1 := \phi(\pi_0)$ verkleinern, so daß (P_{ε_1}) noch lösbar ist mit $\zeta_1^* \geq 0$.

Beweis:

π_0 genügt den Ungleichungsrestriktionen von (P_{ε_1}) und liegt auf mindestens einer der Hyperebenen $H_m^+(\varepsilon_1)$ oder $H_m^-(\varepsilon_1)$.

q.e.d.

IV.12. Algorithmus: (für (Plog))

- (1) Setze $n = 0$. Wähle einen Anfangswert für ϵ_n .

$$\text{Setze } \epsilon_{\max, n} := \epsilon_n$$

$$\text{und } \epsilon_{\min, n} := 0.$$

- (2) Falls $\epsilon_n, n \in \mathbb{N}_0$, bereits klein genug ist, dann ist man fertig. STOP.

Falls ϵ_n größer als ein vorgegebener Wert $\bar{\epsilon}$ (Abbruchkriterium) ist, löse man das Hilfsproblem $(P\epsilon_n)$.

- (3) a) Falls $\zeta_n > 0$, dann kann man ϵ weiter verkleinern:

$$\text{Setze } \epsilon_{n+1} := \epsilon_n - \frac{\epsilon_n - \epsilon_{\min, n}}{2},$$

$$\epsilon_{\max, n+1} := \epsilon_n,$$

$$\epsilon_{\min, n+1} := \epsilon_{\min, n}.$$

Setze $n := n+1$ und springe nach (2).

- b) Falls $\zeta_n < 0$, dann muß ϵ vergrößert werden:

$$\text{Setze } \epsilon_{n+1} := \epsilon_n + \frac{\epsilon_{\max, n} - \epsilon_n}{2},$$

$$\epsilon_{\max, n+1} := \epsilon_{\max, n},$$

$$\epsilon_{\min, n+1} := \epsilon_n.$$

Setze $n := n+1$ und springe nach (2).

- c) Falls $\zeta_n = 0$ (oder $|\zeta|$ klein genug), dann ist man fertig. STOP.

KAPITEL V:

ANWENDUNG DER MATHEMATISCHEN METHODEN AUF DAS

VERKEHRSVERTEILUNGSMODELL

§1. Vorbemerkungen

Die in den voranstehenden Kapiteln angegebenen Verfahren wurden größtenteils programmiert und an der Rechenanlage TR 440 des Regionalen Rechenzentrums Erlangen gerechnet.

Die verwendeten Daten stammen aus dem Generalverkehrsplan (GVP) für Erlangen, der von der Firma Dorsch Consult erstellt wurde ([48]). Unter anderem findet man dort die Matrizen T der Pendler (Individualverkehr) und X der individuellen Reisezeiten zwischen den einzelnen Stadtzonen. Die Stadt wurde dazu in 23 Zonen eingeteilt. Teilweise haben wir diese Daten für ein einfacheres 9-Zonen-Modell aggregiert, indem wir die entsprechenden Pendlerdaten addiert und für die Reisezeiten das arithmetische Mittel gebildet haben. Jede Stadtzone wurde gleichzeitig als Quell- und Zielzone aufgefaßt, so daß

$$I = J = N$$

mit $N = 9$ oder 23 gilt.

Als Verkehrsverteilungsmodell verwendeten wir den Ansatz

$$T_{ij} = e^{\lambda_i - \lambda_j} x_{ij}^\alpha e^{\beta x_{ij}}, \quad 1 \leq i, j \leq N, \quad (V.1.1)$$

wobei T_{ij} die Zahl der Pendler und x_{ij} die Individualreiseden zwischen Zone i und Zone j darstellt.

Für Unimodalitätsuntersuchungen benutzten wir das in der

Praxis nicht sehr gebräuchliche Modell

$$T_{ij} = \hat{T}_{\max} e^{\beta x_{ij}}, \quad (V.1.2)$$

mit $\hat{T}_{\max} = \max_{i,j} \{ \hat{T}_{ij} \}.$

§2. Unimodalitätsuntersuchungen

Diese wurden im einparametrischen Fall am Modell (V.1.2) für die l_1 - und l_2 -Norm mit Sturmschen Ketten und Residuensatz durchgeführt. Bei der l_1 -Approximation erhält man als zu minimierendes Gütekriterium

$$\Phi_1(\beta) := \hat{T}_{\max} \sum_{i=1}^I \sum_{j=1}^J |e^{\beta x_{ij}} - \hat{t}_{ij}|, \quad (V.2.1)$$

bei der l_2 -Approximation

$$\Phi_2(\beta) := \frac{1}{2} \hat{T}_{\max} \sum_{i=1}^I \sum_{j=1}^J (e^{\beta x_{ij}} - \hat{t}_{ij})^2, \quad (V.2.2)$$

wobei jeweils $\hat{t}_{ij} = \hat{T}_{ij} / \hat{T}_{\max}.$

In beiden Fällen numerieren wir die Punkte $(x_{ij}, \hat{t}_{ij})$ so auf einen Index um, daß gilt

$$\frac{\ln \hat{t}_m}{x_m} \leq \frac{\ln \hat{t}_{m+1}}{x_{m+1}}$$

Für $N=23$ folgt dann $M=529$, so daß die Funktion Φ_1 also 529 Stellen hat, an denen sie nicht differenzierbar ist. Folglich müssen 528 Intervalle auf Monotonie der über ihnen definierten Ableitung Φ_1' untersucht werden.

Verwendet man als Einheit für die Reisezeit x_{ij} Minuten (im GVP wird x_{ij} in Sekunden angegeben), dann gilt

$$1 \leq x_{ij} \leq 21, x_{ij} \in \mathbb{N}.$$

Formt man die Ableitungen der Gütekriterien ϕ_1 und ϕ_2 in der in Kapitel II und III beschriebenen Weise durch die Transformation

$$\lambda = e^{\beta \Delta x}$$

mit dem größten gemeinsamen Teiler

$$\Delta x = 1$$

in Polynome um, dann haben diese im Fall der l_1 -Approximation den Grad 19 und bei der l_2 -Approximation den Grad 40.

Insgesamt stellt sich also das numerische Problem, entweder (l_1) 528 mal je ein Problem 19. Grades auf die Zahl seiner Nullstellen in dem Intervall $[e^{\beta_1}, e^{\beta_{i+1}}]$ oder (l_2) ein Polynom 40. Grades einmal über dem Intervall $[e^{\beta_1}, e^{\beta_M}]$ auf Nullstellen zu untersuchen. Wir reduzierten allerdings die Zahl der Intervalle im ersten Problem auf 188, da die übrigen vernachlässigbar klein waren ($\beta_{m+1} - \beta_m < 10^{-3}$).

Bei der Berechnung ergab sich, daß der Zeitaufwand bei Verwendung der Sturmschen Ketten und des Residuensatzes etwa gleich groß ist, nämlich für die Untersuchung des l_1 -Gütekriteriums etwa 15 Minuten und für die des l_2 -Gütekriteriums etwa 5 Sekunden, wobei der große Zeitunterschied auf die verschiedene Zahl von Intervallen (188 bzw. 1) zurückzuführen ist.

Beim Residuensatz wurden die Integrale mit Hilfe des Eulerverfahrens zur numerischen Integration ausgewertet. Um dabei eine Abschätzung für die Rechengenauigkeit zu bekommen, wurden Ober- und Untersummen gleichzeitig berechnet. Eine weitere

Möglichkeit zur Überprüfung der Rechengenauigkeit bietet sich dadurch, daß die Imaginärteile der Integrale in (II.1.24) und (III.1.6) natürliche Zahlen sein und die Realteile verschwinden müssen. Die Kontrolle der Rechengenauigkeit ist deswegen notwendig, weil sich aufgrund der Rundungsfehler falsche Entscheidungen ergeben können, etwa wenn das Integral den Wert 0.5 annimmt. Anwendung der Unimodalitätskriterien aus Residuensatz und Sturmschen Ketten ergab, daß die l_1 - und l_2 -Gütekriterien für (V.1.2) unimodale Funktionen von β sind.

§3. Optimierungsverfahren für das l_2 -Kriterium

Zu minimieren ist die Funktion

$$\phi(\pi) := \sum_{i=1}^N \sum_{j=1}^N (e^{\lambda_i - x_j} x_{ij}^\alpha e^{\beta x_{ij}} - \hat{t}_{ij})^2 \quad (\text{V.3.1})$$

mit $\pi := (\lambda_1, \dots, \lambda_N, x_1, \dots, x_N, \alpha, \beta)^t =: (\lambda^t, \beta^t)^t$,
 $\lambda := (\lambda_1, \dots, \lambda_N, x_1, \dots, x_N)^t$,
 $\beta := (\alpha, \beta)^t$.

Das Optimierungsproblem schreiben wir nun kürzer

$$(P2): \min_{\lambda, \beta} \phi(\lambda, \beta) \quad (\text{V.3.2})$$

und lösen es mittels Blockiteration durch sukzessive Behandlung der beiden Teilprobleme

$$(PA): \min_{\lambda} \phi(\lambda, \beta) \quad (\text{V.3.3})$$

und $(P\beta): \min_{\beta} \phi(\lambda, \beta). \quad (\text{V.3.4})$

Zur Minimierung verwenden wir zwei verschiedene Verfahren, die miteinander schneller gegen das Minimum des Gesamtproblems konvergieren, als wenn man die Teilverfahren auf das Problem (P2) ausweiten würde. Das wäre überdies bei dem be-

schriebenen Fixpunktverfahren gar nicht möglich, da es wesentlich darauf beruht, daß bei den unbekannten Potentialen λ_1 und λ_j nur die Faktoren +1 und -1 stehen, was bei den Parametern α und β nicht zutrifft. Wollte man andererseits das modifizierte Newtonverfahren außer für die Parameter auch noch für die Potentiale benutzen, so müßte man hier statt einer (2,2)- eine (47,47)-Hessematrix aufstellen und eventuell invertieren.

Auf (P λ) wenden wir den Algorithmus III.15, also ein Fixpunktverfahren an, auf (P β) ein modifiziertes Newtonverfahren mit konstanter Schrittweite, das wir im folgenden kurz beschreiben wollen.

Wir betrachten dazu ϕ nur als Funktion von β und schreiben

$$\phi = \phi(\beta), \quad (\text{V.3.5})$$

$$\text{grad } \phi = \text{grad } \phi(\beta) \quad (\text{Gradient von } \phi \text{ bzgl. } \beta),$$

$$H = H(\beta) \quad (\text{Hessematrix von } \phi \text{ bzgl. } \beta).$$

Die Gleichung des Newtonverfahrens zur Minimierung der Funktion ϕ lautet dann

$$\beta^{(n+1)} = \beta^{(n)} + \Delta \beta^{(n)} \quad (\text{V.3.6})$$

$$\text{mit } \Delta \beta^{(n)} = h_n (H(\beta^{(n)}))^{-1} \text{grad } \phi(\beta^{(n)}) \quad (\text{V.3.7})$$

und h_n = Schrittweitenfaktor mit Vorzeichen.

Bekanntlich muß, damit das Newtonverfahren gegen ein Minimum strebt, die Hessematrix positiv definit sein. Da dies in unserem Fall sicher nicht für alle β zutrifft, und man sich anhand der graphischen Darstellungen (z.B. Abb.22) leicht überzeugt, daß auch Bereiche negativer Definitheit und Indefinitheit auftreten, muß man am Newtonverfahren Modifikationen vornehmen.

Nach der Cramerschen Regel zur Berechnung der Inversen einer regulären Matrix gilt:

$$(H(\beta^{(n)}))^{-1} = \frac{\text{adj}(H(\beta^{(n)}))}{\det(H(\beta^{(n)}))}, \quad (V.3.8)$$

wobei $\text{adj}(A)$ die Adjunkte zur Matrix A bezeichnet.

Dividiert man in (V.3.8) statt durch die Determinante durch deren Absolutbetrag, dann wird $\Delta\beta^{(n)}$ von der Definitheit der Hessematrix unabhängig. Damit die Konvergenz gegen ein Minimum von ϕ erhalten bleibt, muß man allerdings nach dem ersten Schritt des Algorithmus einmal feststellen, ob das Verfahren in Richtung fallender Funktionswerte läuft und widrigenfalls das Vorzeichen von h_n in Gleichung (V.3.6) ändern.

Eine zweite Modifikation des Newtonverfahrens wird auf Grund der großen Absolutbeträge, die von ϕ und den Komponenten von $\text{grad}\phi$ und H angenommen werden können ($\|\text{grad}\phi\|_2 \approx 10^9$, $\|\text{adj}(H)\text{grad}\phi\|_2 \approx 10^{16}$), nötig. Um in diesem Fall übersgroße Schrittweiten, die zur Divergenz des Verfahrens führen könnten, zu vermeiden, normieren wir $\frac{1}{h_n} \Delta\beta^{(n)}$ auf 1, was außerdem den Vorteil hat, daß man nun nicht mehr die Determinante von $H(\beta^{(n)})$ benötigt. Insgesamt verwenden wir also das folgende Abstiegsverfahren:

$$\beta^{(n+1)} = \beta^{(n)} + \Delta\beta^{(n)} \quad (V.3.9)$$

mit

$$\Delta\beta^{(n)} := h_n \frac{\text{adj}(H(\beta^{(n)})) \cdot \text{grad}\phi(\beta^{(n)})}{\|\text{adj}(H(\beta^{(n)})) \cdot \text{grad}\phi(\beta^{(n)})\|_2}. \quad (V.3.10)$$

h_n enthält dabei wieder das Vorzeichen und einen Schrittweitenfaktor.

Im allgemeinen wird man bei der numerischen Behandlung auf

Grund der konstanten Schrittweite das Minimum nicht direkt erreichen, sondern daran vorbei- oder darüber hinauslaufen. Daß beispielsweise zwischen $\beta^{(n)}$ und $\beta^{(n+1)}$ ein Punkt β^* mit $\text{grad } \phi(\beta^*) = 0$ vorliegt, erkennt man daran, daß die Gradientenvektoren in diesen beiden Punkten (wenn der endlichen Schrittweite allerdings nur näherungsweise) antiparallel sind. Dies macht man sich folgendermaßen klar:

Sei $h_n = \pm \Delta t$. Dann folgt aus (V.3.7)

$$\Delta \beta^{(n)} = \pm \Delta t (H(\beta^{(n)}))^{-1} \text{grad } \phi(\beta^{(n)}) \quad (\text{V.3.11})$$

und im Grenzwert für $t \rightarrow 0$

$$H(\beta) \frac{d\beta}{dt} = \pm \text{grad } \phi(\beta) \quad (\text{V.3.12})$$

bzw.

$$\frac{d}{d\beta}(\text{grad } \phi(\beta)) \frac{d\beta}{dt} = \pm \text{grad } \phi(\beta) \quad (\text{V.3.13})$$

Diese Differentialgleichung läßt sich explizit lösen:

$$\text{grad } \phi(\beta(t)) = \text{grad } \phi(\beta(0)) e^{\pm t}, \quad (\text{V.3.14})$$

Das bedeutet, daß Newtonverfahren zur Minimierung für hinreichend kleine Schrittweite sukzessiv Punkte berechnen, in denen der Gradient der zu minimierenden Funktion parallel zum Gradienten des Ausgangspunktes ist.

Wenn das Vorzeichen von h_n so gewählt ist, daß

$$\phi(\beta^{(n)}) \leq \phi(\beta^{(n+1)})$$

gilt, dann kann ein Punkt β^* mit

$$\phi(\beta^*) = 0$$

ein Minimum oder ein Sattelpunkt sein, was man durch Untersuchung der Hessematrix $H(\beta^*)$ entscheidet. Falls ein Sattelpunkt vorliegt und $\beta^{(n)}$ und $\beta^{(n+1)}$ Punkte "vor" bzw. "hinter" β^* sind, so muß man im Punkt $\beta^{(n+1)}$ das Vorzeichen von h_n umkehren, um von β^* weg und weiter in Richtung eines Minimums zu laufen.

Bemerkung:

Eine Erweiterung des beschriebenen modifizierten Newtonverfahrens ist die Trajektorienmethode von F. Branin ([9][2][42]) zur globalen Minimierung. Dort wird von der Differentialgeometrie (V.3.13) ausgegangen. Mittels numerischer Integration, die die Gleichungen ((V.3.9) und (V.3.10)) liefert, wird eine Trajektorie berechnet, die Punkte mit Gradienten verbindet, die zu dem Gradienten im Ausgangspunkt parallel sind. Dabei werden wegen (V.3.14) auch Punkte erreicht, in denen der Gradient verschwindet, wobei jeweils anhand der Hessematrix entschieden wird, ob es sich um Maximum, Minimum oder Sattelpunkt handelt, und dann das Vorzeichen von h_p geändert. Das hat zur Folge, daß aus dem anziehenden ein abstoßender "kritischer Punkt" (mit $\text{grad} \phi(\beta^*) = 0$) wird und man zum nächsten Punkt mit Nullgradient weiterläuft. Allerdings kann nicht garantiert werden, daß auf diese Weise alle Punkte mit Nullgradient gefunden werden ([42]).

Bei der Anwendung der Braninschen Trajektorienmethode auf $(P|\beta)$ stellten wir für die von uns benutzten Daten jeweils nur einen Punkt β^* mit $\text{grad} \phi(\beta^*) = 0$ fest. Es erwies sich jedoch als problematisch zu entscheiden, wann die Trajektorie divergiert, d.h. wann eine der Komponenten von β gegen ∞ strebt.

So stellte sich bei einem Testbeispiel mit fünf zu approximierenden Wertepaaren $(x_m, \hat{y}_m)$ und dem Gütekriterium

$$\phi(\alpha, \beta) = \frac{1}{2} \sum_{m=1}^5 (x_m^\alpha e^{\beta x_m} - \hat{y}_m)^2,$$

das ein relatives Maximum, drei relative Minima und drei Sattelpunkte besitzt, heraus, daß die Trajektorien je nach Start-

wert (α_0, β_0) teilweise weit (im Vergleich zum Abstand der kritischen Punkte untereinander) von den kritischen Punkten weglief, um dann doch wieder zurückzukehren.

Auch erwies es sich dabei als sinnvoll, von Startwerten mit solchen Gradienten auszugehen, deren Komponenten annähernd gleich groß sind, um so numerische Probleme zu vermeiden, die sonst in der Umgebung kritischer Punkte auftreten, wo relativ kleine Änderungen von α und β große Änderungen des Gradienten und damit auch seiner Komponenten zueinander nach sich ziehen können.

Um Startwerte für das Gesamtverfahren (V.3.2) zu bekommen, wurde zunächst mittels der Fehlerquadratmethode das überbestimmte lineare Gleichungssystem

$$\lambda_i - x_j + \alpha \ln x_{ij} + \beta x_{ij} = \ln \hat{T}_{ij} \quad (\text{V.3.15})$$

gelöst (vgl. Kap. III, §4).

Die Tabelle auf der folgenden Seite zeigt unter anderem die Verbesserung der Approximationsgüte, die man erreicht, wenn man die Gleichungen gemäß den Überlegungen von Kapitel III, §4 gewichtet.

Um Vergleiche mit einem Verfahren anstellen zu können, das die Parameter so bestimmt, daß die Konsistenzbedingungen (I.3.6) und (I.3.7) erfüllt sind, wurde das Fixpunktverfahren zur Bestimmung der Potentiale gegen ein Knotenanalyseverfahren (Algorithmus I.9) ausgetauscht, das modifizierte Newtonverfahren zur Berechnung der optimalen Parameter jedoch beibehalten. Dabei zeigte sich im Vergleich zu dem von uns verwendeten Verfahren, das den Konsistenzbedingungen nur näherungsweise

(Forts. übernächste Seite)

Vergleich verwendeter l_2 -Approximationsmethoden (N=23)

Gütekriterium	$(\sum_{i=1}^n (T_{ij} - \hat{T}_{ij})^2)^{1/2}$	Knoten- anzahl (sec)	$\sum_{i=1}^n T_{ij} - \hat{T}_{ij} $	$\sum_{i=1}^n T_{ij} - \hat{T}_{ij} $	$\sum_{i=1}^n T_{ij} - \hat{T}_{ij} $	$\sum_{i=1}^n T_{ij} - \hat{T}_{ij} $
...	l_2		l_1	l_∞		
Fehlerquadrat- methode un- gewichtet	15401 (11.03)	35	70818 (50.0) 25.4%	12412		159036 (113.0)
Fehlerquadrat- methode gewichtet	1281 (3.07)	25	78316 (56.07) 28.0%	825		190541 (142.9)
Koordinaten- abstiege	3296 (2.36)	360	55843 (39.98) 20.0%	259		151224 (108.3)
Fixpunkt und modifizierter Newton	3150 (2.26)	95	49923 (35.74) 17.9%	843		146211 (104.7)
Knotenanalyse und modifi- zierter Newton	3652 (2.61)	200	47230 (33.81) 16.9%	1396		139678 (100.0)

Die obere Zahl stellt jeweils den Absolutwert des Gütekriteriums und die in Klammern gesetzte untere der durch die empirische Gesamtänderungszahl (139678) dividierten Relativwert in Prozent dar. Die dritte Zahl gibt die Zahl der festverteilten Pendler in Prozent wieder.

"Fixpunkt und modifizierter Newton" teilt für das Blockiterationenverfahren aus Algorithmus III.15 für λ und dem auf S.100 ff beschriebenen modifizierten Newtonverfahren, "Knotenanalyse und modifizierter Newton" für die auf S.104 beschriebene Methode.

Obwohl bei sämtlichen in der ersten Spalte der Tabelle angegebenen Verfahren das l_2 -Kriterium zu optimieren war - die Werte des Gütekriteriums finden sich in der zweiten Spalte - geben wir zum Vergleich in weiteren Spalten die Werte des l_1 - und des l_∞ -Kriteriums im l_2 -Optimum an, sowie die Zahl der insgesamt verteilten Pendler.

genügt, eine Verdopplung des Rechenzeitaufwandes für die Bestimmung der optimalen Parameter und Potentiale. Zwar werden dort die Konsistenzbedingungen exakt erfüllt, die Approximationsgüte (V.3.1) ist jedoch schlechter, da in sie nur die Pendlerzähler auf den (ij) -Bögen und nicht die Erfüllung der Konsistenzbedingungen eingehen. Betrachtet man allerdings für unser Verfahren die Knoten, in denen die Konsistenzbedingungen schlecht erfüllt sind, in denen also

$$\left| \sum_{j=1}^J T_{ij} - \sum_{j=1}^J \hat{T}_{ij} \right| \text{ oder } \left| \sum_{i=1}^I T_{ij} - \sum_{i=1}^I \hat{T}_{ij} \right|$$

große Absolutwerte annehmen, und vergleicht diese mit den Knoten, in denen bei Erfüllung der Konsistenzbedingungen durch das andere Verfahren die Werte

$$\sum_{j=1}^J |T_{ij} - \hat{T}_{ij}| \text{ oder } \sum_{i=1}^I |T_{ij} - \hat{T}_{ij}|$$

groß werden, in denen also die Zahlen der ausgehenden oder einmündenden Fahrten schlecht approximiert werden, so stellt man fest, daß es sich im wesentlichen um dieselben Knoten handelt. Hier scheinen entweder empirische Daten mit großen Meßfehlern vorzuliegen oder das Modell die Wirklichkeit nicht gut wiederzugeben.

Insgesamt scheint daher keine der beiden Vorgehensweisen bei der Kalibrierung der Parameter der anderen gegenüber deutliche Vor- oder Nachteile zu haben, so daß dem Benutzer die Entscheidung überlassen bleibt, ob er ein schnelleres Verfahren vorzieht, das die Pendlerdaten besser approximiert, oder ein langsames, das dafür die Konsistenzbedingungen erfüllt.

4. Optimierungsverfahren für das 1- Gütekriterium

Das Optimierungsproblem $(P\mathcal{E}_0)$ aus Kapitel IV führt zu folgendem linearen Programm.

Lineares Programm:

$$\text{Max } \zeta = \text{Max } (\zeta, \pi^t) (1, \theta^t)^t$$

unter den Nebenbedingungen

$$(\zeta, \pi^t) (1, A) \leq e^t$$

mit freien Variablen ζ, π .

Dabei bedeuten

$$1 := (1, \dots, 1)^t \in \mathbb{R}^{2N+2},$$

$$e := (\ln(p_{11} + \varepsilon), \dots, \ln(p_{NN} + \varepsilon), -\ln^*(p_{11} - \varepsilon), \dots, -\ln^*(p_{NN} - \varepsilon))^t,$$

die Inhomogenität des linearen Programms,

$$\theta := (0, \dots, 0)^t \in \mathbb{R}^{2N+1},$$

$$A := \text{Systemmatrix vom Typ } (2N^2, 2N+1).$$

Die Matrix A hat folgende Blockstruktur:

$$A = \begin{pmatrix} D & LX & X \\ -0 & -LX & -X \end{pmatrix}, \text{ wobei}$$

$$D^t := \left(\begin{array}{ccccccccc} 1 & \dots & 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & : & \dots & : \\ \vdots & & : & 0 & \dots & 0 & & : & & : \\ & & : & \vdots & & \vdots & & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \\ -1 & & & -1 & & & \dots & -1 & & \\ & \cdot & & & \cdot & & & \cdot & & \\ & & \cdot & & & \cdot & & & \cdot & \\ & & & -1 & 0 & & -1 & 0 & \dots & -1 & 0 \end{array} \right) \left. \begin{array}{l} \text{N Zeilen} \\ \text{N-1 Zeilen} \end{array} \right\}$$

N N N Spalten

$$LX := (\ln x_{11}, \dots, \ln x_{1N}, \dots, \ln x_{N1}, \dots, \ln x_{NN})^t,$$

$$X := (x_{11}, \dots, x_{1N}, \dots, x_{N1}, \dots, x_{NN})^t.$$

D^+ ist die um die letzte Zeile verminderte Knoteninzidenzmatrix des Graphen, der entsteht, wenn man in dem in I.2.1 dargestellten Netz die Quell- und Zielbögen wegläßt.

Da A im allgemeinen mehr Spalten als Zeilen hat ($2N^2$ - gegenüber $2N+1$), ist es sinnvoll, das zugehörige duale Programm zu lösen.

Duales Programm:

$$\text{Min } c^t z$$

unter den Nebenbedingungen

$$(1 \ A)^t z = \begin{pmatrix} 1 \\ e \end{pmatrix}$$

und $z \geq 0$ (also keine freie Variable), $z \in \mathbb{R}^{2N+2}$.

Für den Zusammenhang zwischen Lösbarkeit des dualen Programms und Verträglichkeit der Nebenbedingungen des primalen Programms gelten folgende Standardaussagen der linearen Programmierung [46] .

1. Lösbarkeit des dualen Programms

impliziert Lösbarkeit des primalen Programms, und diese die Verträglichkeit der Nebenbedingungen.

2. Nichtlösbarkeit des dualen Programms (keine zulässige Ausgangslösung)

impliziert Unbeschränktheit des primalen Programms.

3. Unbeschränktheit des dualen Programms

impliziert Nichtlösbarkeit des primalen Programms, und diese die Unverträglichkeit der Nebenbedingungen.

Bemerkungen:

1. Die Unbeschränktheit des dualen Algorithmus ist dadurch gekennzeichnet, daß der reduzierte Kostenkoeffizient mindestens einer Spalte des Tableaus negativ ist, man aber kein Pivotelement finden kann, da alle Elemente der Spalte negativ oder gleich Null sind (vgl. [46], S.34).
2. Falls in einer der Komponenten von c gilt $\hat{T}_{ij} - \varepsilon < 0$ und damit $\ln^*(\hat{T}_{ij} - \varepsilon) = -\infty$, so genügt es, als Wert des Logarithmus eine hohe negative Zahl zu setzen (z.B. -10^{-6}), um die entsprechende Bedingung zu deaktivieren.
3. Der Anfangswert von ε sollte kleiner als der größte $\hat{T}_{ij}$ -Wert sein, damit wenigstens eine der Bedingungen

$$-\xi + \lambda_i - \alpha_j + \alpha \ln x_{ij} + \beta x_{ij} = \ln^*(\hat{T}_{ij} - \varepsilon)$$
 (Zeilenindex im primalen Programm $> N^2$) aktiviert ist. Meist hat man dann schon Lösbarkeit des dualen linearen Programms.
4. Unter der in 3. getroffenen Voraussetzung an ε erhält man sofort eine, wenn auch degenerierte zulässige Lösung des dualen linearen Programms, indem man den Vektor $(1, \theta^t)^t$ wie folgt zusammensetzt:
 Sei $\hat{T}_{lk} = \max T_{ij}$. Die beiden zugehörigen Nebenbedingungen des primalen linearen Programms stehen dann als Spalten mit den Indices $(l-1)N+k$ und $(l-1)N+k+N^2$ in der Koeffizientenmatrix des dualen linearen Programms. Diese beiden Spalten haben in der ersten Komponente den Wert 1 und sind in den übrigen Komponenten entgegengesetzt gleich, so daß ihre mit 0.5 gewichtete Summe gerade die Inhomogenität $(1, \theta^t)^t$ des dualen linearen Programms ergibt.

Bezeichnungen

		Seite
A	Parametervektor bei Exponentialsummen	vi
B	Menge der Bögen	4
ρ	Parametervektor	13
β_m	Parameter	i
C	Gesamtreisekosten	9
c_{ij}	verallgemeinerte Reisekosten von i nach j	ii
D_j	Gesamtzahl aller in Zone j einmündenden Fahrten	i
D	Knoteninzidenzmatrix	4
exp	Exponentialfunktion	
Φ	Gesamtfehlerfunktion (je Kapitel verschieden)	
φ	Einzelfehlerfunktion (" ")	
g_{ij}	Gewichtung	67
I	Zahl der Quellzonen	
i	Quellzonenindex (gelegentlich auch imaginäre Einheit, Verwechslungen sind nicht zu befürchten)	
J	Zahl der Zielzonen	
j	Zielzonenindex	
$\mathcal{K}$	Kennlinie	6
k_{ij}	die Kennlinie definierende Funktion	6
K	Menge der Knoten	4
k_j	Zielzonenparameter	i
$\mathbf{k}$	Vektor der Zielzonenparameter	67
π_j	Zielzonenpotential	4
ln	natürlicher Logarithmus	
ln	verallgemeinerter natürlicher Logarithmus	89
$\mathcal{L}$	Likelihood-Funktion	27
L	log-Likelihood-Funktion	27
l_i	Quellzonenparameter	i
$\mathbf{l}$	Vektor der Quellzonenparameter	67
$L(A, x)$	Exponentialsumme	vi
λ	Quellzonenpotential	4
$\boldsymbol{\lambda}$	Vektor der Potentiale	16
N	Zonenzahl	
$\mathbb{N}$	Menge der natürlichen Zahlen	

π	Parametervektor	21
P1	l_1 -Approximation	30
P2	l_2 -Approximation	56
P2a	l_2 -Approximation Spezialfall	66
P1n	$\log$ - l_2 -Approximation	76
P_{ω}	l_{ω} -Approximation	84
Pexp	umformulierte l_{ω} -Approximation	90
Plog	umformulierte l_{ω} -Approximation	90
$P\varepsilon_0$	Hilfsproblem für festen ε_0	93
$P\lambda$	Fixpunktiteration	99
$p\beta$	modifiziertes Newtonverfahren	99
$\mathbb{R}$	Menge der reellen Zahlen	
$\mathbb{R}_+^N$	positiver Orthant des $\mathbb{R}^N$	
S, S^*	Entropie	11
T	Gesamtzahl aller Fahrten	8
$T_{i,j}$	Fahrten von i nach j	i
$\mathbf{t}$	Vektor der $T_{i,j}$ (in lexikographischer Reihenfolge)	
$\mathbf{T}$	Matrix der $T_{i,j}$	
$\mathbf{x}$	Vektor der erklärenden Variablen	
$x^{(n)}, x^m$	Komponente von $\mathbf{x}$	
D^t	zu D transponierte Matrix	
$\mathbf{t}^t$	zu $\mathbf{t}$ transponierter Vektor	
"^"	empirische Größe	
conv	konvexe Hülle	
$\forall$	für alle	

Literatur

- 1 Bacharach, M.:
Biproportional Matrices and Input-Output-Change
Cambridge, University Press (1970)
- 2 Barrodale, I. - Roberts, F.D.K.:
An Efficient Algorithm for Discrete L_1 Linear Approximation
With Linear Constraints
SIAM J. Numer. Anal. 15 (1978), 603-611
- 3 Batty, Michael:
Urban Density and Entropy Functions
Reading Geographical Papers No. 36
Dpt. of Geography, Univ. of Reading, 1975
- 4 Berge, Claude:
Espaces Topologiques - Fonctions Multivoques
Dunod, Paris (1966)
- 5 Box, M.J.:
Bias in Nonlinear Estimation
Royal Statistical Society, Research section, Dec. 9th
(1970)
- 6 Braess, Dietrich:
Über das Anzahlproblem bei der rationalen L_2 -Approximation
in: Numer. Meth. d. Approx. Th. III
(Tagung Oberwolfach 25.-31.5.1975)
Herausgeber: L. Collatz, H. Werner, G. Meinardus
Birkhäuser (1976)
- 7 Braess, Dietrich:
On Rational L_2 -Approximation
J. Approx. Th. 18 (1976), 136-151
- 8 Braess, Dietrich:
Approximation mit Exponentialsummen
Computing 2 (1976)
- 9 Branin, Franklin H. Jr. - Hoo, Stanley K.:
A Method for Finding Multiple Extrema of a Function of
N Variables
Proceedings on the "Conference on Numerical Methods for
Nonlinear Optimisation", Dundee, Scotland (1971)
in: Numerical Methods of Nonlinear Optimisation,
Academic Press, London (1972)
- 10 Cesario, Frank J.:
Least-Squares Estimation of Trib Distribution Parameters
Transn. Res. 9 (1975), 13-18
- 11 Dhrymes, Phoebus J.:
On Deriving Unbiased Estimators for the Parameters of
the Cobb-Douglas-Production Function
Econometrica 30 (1962), 297-304

- 12 Dixon, L.G.W. - Szecő, G.P.:
Towards Global Optimisation
(Workshop Cagliari, Oct. 1974)
American Elsevier - N.Y. (1975)
- 13 Donath, G. - Elster, K.H.:
Generalized Unimodal Functions
Colloquia Mathematica Societatis János Bolyai
12. congress in OR, Eger (Hungary) (1974)
in: "Progress in OR" Volume I
Prékopa A.
North Holland, Amsterdam 1976
- 14 Dunham, Charles R.:
Chebyshev Approximation by Exponential-Polynomial
Products
J. Approx. Th. 14 (1975), 77-78
- 15 Erlander, S. = Stewart, N.F.:
Interactivity, Accessibility and Cost Trip Distribution
Transpn. Res. 12 (1978), 291-293
- 16 Evans, Andrew W.:
Some Properties of Trip Distribution Methods
Transpn. Res. 4 (1970), 19-36
- 17 Evans, Andrew W.:
The Calibration of Trip Distribution Models with
Exponential or Similar Cost Functions
Transpn. Res. 5 (1971), 15-38
- 18 Evans, Suzanne P.:
A Relationship between the Gravity Model for Trip
Distribution and the Transportation Problem in Linear
Programming
Transpn. Res. 7 (1973), 39-61
- 19 Evans, Suzanne P. - Kirby, Howard R.:
A Three-dimensional Furness Procedure for Calibrating
Gravity Models
Transpn. Res. 8 (1974), 105-122
- 20 Feynman, R.P.:
Statistical Mechanics - A Set of Lectures
W.A. Benjamin Inc., Advanced Book Program
Reading, Mass.
- 21 Gantmacher, F.R.:
Matrizenrechnung II
VEB Deutscher Verlag der Wissenschaften, Berlin (1966)
- 22 Goldberger, Arthur S.:
The Interpretation and Estimation of the Cobb-Douglas-
Functions
Econometrica 35 (1968), 464-472

- 23 Goldfeld, Stephen M. - Quandt, Richard E.:
Nonlinear Methods in Econometrics
North Holland, Amsterdam - London 1972
- 24 Hansen, Stein:
Entropy and Utility in Traffic Modelling
in: "Transportation and Traffic Theory"
Proceedings of the 6th International Symposium on Transportation and Traffic Theory, Sidney 1974
D.J. Buckley ed.
Elsevier, N.Y. - London - Amsterdam, 435-482
- 25 Intriligator, Michael D.:
Econometric Models, Techniques, and Applications
North Holland, Amsterdam - Oxford
Prentice Hall, Englewood Cliffs, N.J. 1978
- 26 Jacobs, Konrad:
Elemente der Informationstheorie
Skriptum
- 27 John, Fritz:
Lectures on Advanced Numerical Analysis
Gordon and Breach, N.Y. - London - Paris, 1967
- 28 Kammler, David K.:
Existence of Best Approximations by Exponential Sums in
Several Independent Variables
J. Approx. Th. 11 (1974), 143-148
- 29 Kammler, David K.:
Approximation with Sums of Exponentials in $L_p[0, \infty)$
J. Approx. Th. 16 (1976), 348-408
- 30 Kammler, David K.:
Existence of Good and Best Approximations on Unbounded
Domains by Exponential Sums in Several Independent
Variables
J. Approx. Th. 21 (1977), 215-223
- 31 Kirby, Howard R.:
Theoretical Requirements for Calibrating Gravity Models
Transpn. Res. 8 (1974), 97-104
- 32 Meinardus, G. - Schwedt, D.:
Nichtlineare Approximationen
Arch. Rational Mech. Anal. 17 (1964), 297-326
- 33 Meulenberg, M.T.G.:
On the Estimation of an Exponential Function
Econometrica 33 (1965), 863-868
- 34 Murchland:
Some Remarks on the Gravity Model of Traffic Distribution
and an Equivalent Maximization Formulation
LSE-TNT-38, London School of Economics, London, July 1966

- 35 Polumsky, Dieter:
Zentralitätsaspekte im regionalen Reiseverhalten
Dissertation, Aachen 1976
- 36 Remez, E.:
Sur le calcul effectif des polynoms d'approximation
de Tschebyscheff
C.R. Acad. Sci. Paris 199 (1934), 337-340
- 37 Reilly, W.F.:
The Law of Retail Gravitation
G.P. Putnam & Son N.Y. (1931)
- 38 Rice, John R.:
Chebyshev Approximation by Exponentials
J.SIAM 10 (1962), 149-161
- 39 Rice, John R.:
Algorithms for Chebyshev Approximation by ab^x+c
J.SIAM 9 (1961), 571-583
- 40 Rice, John R.:
Chebyshev Approximation by ab^x+c
J.SIAM 8 (1960), 691-702
- 41 Spieß, J.:
Uniqueness Theorems for Nonlinear L_2 -Approximation Problems
Computing 9 (1973)
- 42 Treccani, G.:
On the Convergence of Branin's Method: a Counterexample
in: "Towards Global Optimisation" ed. Dixon - Szegö
North Holland - American Elsevier 1975
- 43 Van der Vet, R.P.:
Flexible Solutions to Systems of Linear Inequalities
European J. of OR 1 (1977), 247-254
- 44 Wilson, A.G.:
A Statistical Theory of Spatial Distribution Models
Transpn. Res. 1 (1967), 253-270
- 45 Wilson, A.G.:
Advances and Problems in Distribution Modelling
Transpn. Res. 4 (1970), 1-18
- 46 Luenberger, David G.:
Introduction to Linear and Nonlinear Programming
Addison-Wesley Publishing Company, Reading, Mass. (1973)
- 47 Ortega, J.M. - Rheinboldt, W.C.:
Iterative Solution of Nonlinear Equations in Several
Variables
Academic Press, N.Y. - London (1970)
- 48 Dorsch Consult:
Gesamtverkehrsplan Erlangen
März 1977

Lebenslauf

Am 23. Januar 1949 wurde ich als Sohn des kaufmännischen Angestellten Hans Götz und seiner Ehefrau Helene, geb. Holzer, in Marktredwitz (Kreis Wunsiedel im Fichtelgebirge) geboren.

Vom Herbst 1955 bis zum Sommer 1959 besuchte ich die Volksschule in Thiersheim, anschließend das Otto-Hahn-Gymnasium in Marktredwitz, an dem ich im Juli 1968 das Abitur bestand.

Im Wintersemester 1968/69 begann ich an der Universität Erlangen-Nürnberg das Studium der Mathematik und Physik und legte dort im Herbst 1973 das Staatsexamen ab.

Vom 1. Januar 1974 bis zum 31. Dezember 1979 war ich mit einer Unterbrechung vom 1. Januar 1975 bis zum 31. März 1976, während der ich meinen Grundwehrdienst in Erndtebrück und Hof ableistete, als Verwalter der Dienstgeschäfte eines Wissenschaftlichen Assistenten am Institut für Angewandte Mathematik, Lehrstuhl Prof. Dejon, beschäftigt.

Erlangen, Januar 1980

